



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



Publication number: **0 616 291 A2**

(12)

## EUROPEAN PATENT APPLICATION

(21) Application number: 94101489.6

(51) Int. Cl.<sup>5</sup>: G06F 15/80

(22) Date of filing: 01.02.94

(30) Priority: 18.03.93 JP 58328/93

(43) Date of publication of application:  
21.09.94 Bulletin 94/38

(84) Designated Contracting States:  
CH DE FR LI SE

(71) Applicant: HITACHI, LTD.  
6, Kanda Surugadai 4-chome  
Chiyoda-ku, Tokyo 100 (JP)

(72) Inventor: Takahashi, Isao  
Daini Motosoja Jutaku 402,

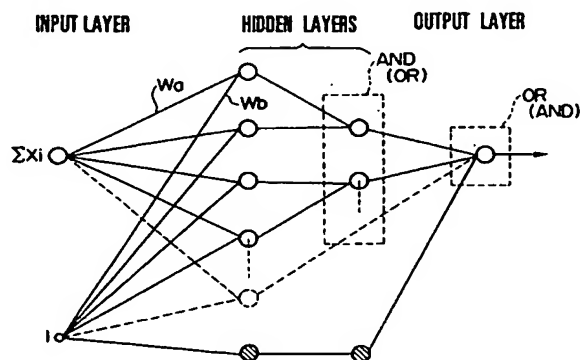
152-3, Motosojamachi  
Maebashi-shi (JP)  
Inventor: Endo, Fumihito  
37-14, Nishinarusawacho-4-chome  
Hitachi-shi (JP)  
Inventor: Yamaguchi, Tokio  
34-22, Onumacho-3-chome  
Hitachi-shi (JP)

(74) Representative: Strehl Schübel-Hopf Groening  
& Partner  
Maximilianstrasse 54  
D-80538 München (DE)

(54) A method of configuring a neural network and a diagnosis/recognition system using the same.

(57) In a recognition/diagnosis method, a relation between total sums of inputs of respective cases and teacher data is listed in the order of magnitude of the total sums of inputs. Based on the value of the teacher data for the input having the maximum total sum and the number of times of change of teacher data in the table are considered, a configuration of a neural network (the number of hidden layers and the number of neurons thereof) is determined. Coupling weights are analytically calculated based on the table.

FIG.2



EP 0 616 291 A2

## BACKGROUND OF THE INVENTION

## FIELD OF THE INVENTION

5 The present invention relates to a recognition or diagnosis method primarily using a neural network for diagnosing the presence or absence of a fault in a current operation status by utilizing past operation information of various components of a power supply circuit such as a gas insulated switchgear (GIS) or a transformer. It is applicable to not only various engineering fields but also a medical field, and more generally to the recognition and diagnosis of physical, chemical and biological phenomena.

10 DESCRIPTION OF THE RELATED ART

In JP-A-2-297074, a diagnosis method by using a neural network is proposed. In the method, past status of various types of equipments are previously learnt by a neural network having a multi-layer structure. When new status of the equipments is inputted, whether the equipments are normal or not is determined based on the learn result. If they are determined to be abnormal, a cause therefor is inferred.

In JP-A-2-272326, detailed description is made on mechanical vibration, acoustic vibration and electric oscillation of a rotary machine.

15 In JP-A-2-100757, a learning method of a parallel neural network is proposed. It is stated therein that a back propagation method is becoming effective as a learning method. A conventional learning method has disadvantages in that the learning does not proceed once it falls in a local minimum and a precision of learning sharply decreases when the number of data to be learnt is large. In the proposed method, more than one neural networks are connected in parallel to serially learn in order to improve an efficiency and the precision of the learning. Other arts related to the present invention are disclosed in USP 5,023,045, JP-A-2-206721, JP-A-4-120479, JP-A-3-201079 and JP-A-3-092975.

20 Fig. 1 shows a conventional neural network. Any number (one in the illustrated example) of hidden layers are provided between an input layer and an output layer. A set of input data  $X_1, X_2, X_3, \dots, X_n$  are inputted to the input layer. In Fig. 1, "1" is also always supplied for adjusting outputs from neurons. Products of those inputs and coupling weights  $W_1, W_2, W_3, \dots, W_n, W_{n+1}$  are inputs  $U_1$  to a neuron Ne. An output from the neuron is  $V_1$ .  $U_1$  and  $V_1$  are calculated by equations (1) and (2).

$$U_1 = \sum_{i=1}^n X_i \cdot W_i + 1 \cdot W_{n+1} \quad (1)$$

35

$$V_1 = \frac{1}{(1 + e^{-u_1})} \quad (2)$$

40

Similar equations are applied to the inputs and the outputs of other neurons of the hidden layers and the output layers. Only one  $N_a$  of the neurons of each hidden layer is coupled with "1" for adjusting the output of a neuron connected thereto which the output of the connected is equal to "1" to provide a predetermined coupling weight.

45 Thus, in the conventional neural network, an operation of learning is essential. The learning means that coupling weights among the neurons are determined through iterative calculations so that a calculated value approximated to data called teacher data which is predetermined for input data is outputted from the output layer when the input data is applied. For example, it is the learning to gradually modify the coupling weights through the iterative calculations, to reduce an error which is defined as 0.5 times of a square sum of differences between the teacher data and the calculated output. If the learning proceeds well, the error decreases, but in some cases the error does not decrease and the calculated output is not attained with a sufficient precision even after a large number of times of learning. Such a case is called that the neural network has fallen in a local minimum. In this case, it is not possible to get out of the local minimum whatever number of times of the learning is increased and appropriate coupling weights are not determined.

50 Even if the calculated output converges to a desired value, a large number of times of learning requires a very long time (for example, several hours or more) depending on the scale of the neural network and the teacher data.

In the conventional neural network, the optimum number of hidden layers and the number of neurons are determined by a trial and error method. It is inefficient and the improvement thereof has been desired.

# SUMMARY OF THE INVENTION

5

It is an object of the present invention to provide a method having a high generalization ability to solve the above problems.

A method for analytically determining coupling weights of neurons from input data and teacher data has not been known but the inventors recently developed a method for analytically determining coupling  
10 weights of neurons, which is utilized in the present invention.

Where the coupling weight is to be determined by iterative learning, the above method for analytically determining the coupling weights may be utilized. For example, the number of hidden layers, the number of neurons thereof and what is made clear relating to the coupling weight are partially utilized. The following means is further utilized.

15 (1) As will be described in detail hereinafter, characteristics of input data are examined from as wide range as possible, and learning data is classified into several classes. Then, the proposed method is applied to each of the classes.

(2) A range of scatter of the data to be learnt is considered and design is made such that a correct solution is obtained if diagnosis data is within the range of scatter.

20 (3) If a correct solution is determined after the diagnosis, the diagnosis data and the correct solution are made available as learning data and teacher data in the next run.

(4) In case of diagnosis, the inference is set logically more strict.

The present invention relates to a recognition or diagnosis method using a neural network having an input layer, an hidden layer and an output layer in which learning data inputted to the input layer are  
25 classified into a plurality of types in accordance with characteristics of the data, the learning data is inputted for each type, a total sum is determined for each case, corresponding teacher data and the total sum are associated in the order of magnitude for each output of the output layer, and coupling weights are determined based on the teacher data corresponding to the case having a maximum total sum in the above table, the total sums before and after the change of the teacher data and the teacher data.

30 The present invention further relates to a recognition or diagnosis method using a neural network having an input layer, an hidden layer and an output layer in which input data is inputted for each value for each case based on an upper limit and a lower limit of the input data for each case inputted to the input layer, a total sum is determined for each case, corresponding teacher data and the total sum are associated in the order of magnitude for each output of the output layer, and the coupling weight is determined based on the  
35 teacher data corresponding to the case having a maximum total sum in the above table, the total sums before and after the change of the teacher data and the teacher data.

The input data inputted to the input layer preferably has the same coupling weight multiplied to the upper limit and the lower limit.

The present invention further relates to a recognition or diagnosis method using a neural network having  
40 an input layer, an hidden layer and an output layer in which a an input sum is determined by selecting coupling weights such that the input sum determined from a case other than a particular case inputted to the input layer does not fall between two input sums determined from an upper limit and a lower limit of the input data of the particular case, the input sum is inputted as the input data, an input total sum is determined for each case, corresponding teacher data and the input total sums are associated in the order  
45 of magnitude for each output of the output layer, and the coupling weight is determined based on the teacher data corresponding to the case having the maximum input total sum in the above table, the input total sums before and after the change of the teacher data and the teacher data.

The teacher data of the particular case and the teacher data of the case other than the particular case are preferably different from each other.

50 Preferably, the coupling weight is selected such that the second hidden layer functions as an AND circuit and the output layer functions as an OR circuit when there are two hidden layers and the teacher data corresponding to the case having the maximum total sum is 0. When there are two hidden layers and the teacher data corresponding to the case having the maximum total sum is 1, it is preferable to select the coupling weight such that the second hidden layer functions as an OR circuit and the output layer functions  
55 as an AND circuit.

Further, the input data is preferably a reciprocal for each case so that the magnitude relation of the input data is reversed.

Preferably, a back propagation iterative learning method is used to calculate the coupling weights which are principal parts of the neural network.

The present invention further relates to a recognition or diagnosis method using a neural network having an input layer, an hidden layer and an output layer in which learning data is inputted to the input layer for each case, a total sum is calculated for each case, corresponding teacher data and the total sum are associated in the order of magnitude for each output of the output layer, a coupling weight is determined based on the teacher data corresponding to the case having a maximum total sum in the above table, the total sums before and after the change of the teacher data and the teacher data, and the fact that the inputted diagnosis data is out of the range of scatter designated for the learning data is displayed.

Preferably, the teacher data and the total sums are listed in the order of magnitude to associate the teacher data with the total sums in the order of magnitude.

The present invention further relates to a recognition or diagnosis method using a neural network having an input layer, an hidden layer and an output layer in which when diagnosis data is to be inputted and a correct solution to the diagnosis data is known, the diagnosis data and the corresponding correct solution are transferred as a part of learning data after the diagnosis.

The present invention further relates to a recognition or diagnosis method using a neural network having an input layer, an hidden layer and an output layer in which an output to diagnosis data is calculated after the diagnosis data has been inputted, at least one probable learning data is inferred from the output, the learning data is compared with the diagnosis data, and a result is displayed.

Where the inferred learning data has a scatter, it is preferable to compare the diagnosis data with the learning data having the range of scatter and display a result.

Preferably, the first coupling weight to be multiplied to the input data of the input layer is different from other coupling weights.

Further, in accordance with the present invention, the diagnosis data and the learning data are directly compared, whether the diagnosis data and the learning data of a particular case totally match, partially match or totally mismatch is determined, a result is displayed, the diagnosis data and the learning data having a range of scatter are directly compared, whether the diagnosis data is totally in the range, partially in the range or totally not in the range is determined, and a result is displayed.

Further, in accordance with the present invention, simultaneous equations are prepared by using a calculation formula for the input to the neuron based on an upper limit and a lower limit of the input data having a range of scatter, the simultaneous equations are solved to determine a coefficient, and the neural network is utilized based on the coupling weight.

Further, the present invention relates to a recognition or diagnosis method using a neural network having an input layer, an hidden layer and an output layer weighted by coupling weights which has a capability to evaluate teacher data for determining the coupling weights.

Further, in accordance with the present invention, a system has an electrical circuit to be used under a plurality of conditions and the circuit is diagnosed by using a neural network. The plurality of conditions are taken as learning data to determine coupling weights of the neural network.

The present invention is now summarized. A threshold modifying neuron which always outputs "1" is provided for each layer except the output layer, a total sum of inputs is calculated for each case, total sums are listed together with corresponding teacher data in the order of magnitude of the total sum, neurons equal in number to the number of times of change of the teacher data are sequentially arranged as a first hidden layer based on the number N of times of change of the teacher data in the list, and where N is odd, (N-1)/2 neurons are arranged as a second hidden layer, outputs of the first hidden layer are sequentially inputted to the second hidden layer two at a time, outputs thereof are inputted to the output layer together with the remaining one output of the first hidden layer. Where N is even, N/2 neurons are arranged, the outputs of the first hidden layer are sequentially inputted two at a time, the outputs thereof are inputted to the output layer, and the coupling weights are analytically determined based on the teacher data corresponding the case having a maximum input sum of the table and the input sums before and after the change of the teacher data and the teacher data.

Preferably, any numbers  $K_i$  which are not all equal, for example, a random number are multiplied to the inputs before summation instead of simple summation of the inputs for each case.

Further, it is preferable to provide one threshold modifying neuron which always outputs "1" for each layer except the output layer and a total sum of products of the inputs and appropriate coupling weights is determined for each case.

Further, input sums which are products of an upper limit and a lower limit of the input and an appropriate same coupling weight are prepared for the input data of each case based on the upper limit and the lower limit of the scatter of the input data of each case, and the coupling weights are selected such that

the input sum determined from other cases does not fall between two input sums determined based on the upper limit and the lower limit of the inputs of a particular case. Namely, the weights are determined such that the sum of other case does not fall in the range of the first determined case.

Further, input sums which are products of an upper limit and a lower limit of the input and an appropriate same coupling weight are prepared for the input data of each case based on the upper limit and the lower limit of the scatter of the input data of each case, and the coupling weights are selected such that the input sum determined from other cases having teacher data other than the teacher data of a particular case does not fall between two input sums determined based on the upper limit and the lower limit of the inputs of the particular case.

Further, the fact of diagnosis is preferably indicated by any means such as a display screen of video display means or by alarm sound.

The diagnosis data and the learning data are preferably compared for total match, partial match or total mismatch, and a function to classify the result is preferably provided.

A matrix is preferably used as means for comparing data numbers of the diagnosis data and the learning data.

(1) By the use of the analytical approach, the iterative learning required in the prior art is not necessary. An optimum number of hidden layers of the neural network and the number of neurons can be predetermined. No problem of local minimum arises. Accordingly a time from the input of data to the diagnosis by using the neural network can be reduced. Since the present method may be applicable irrespective of the number of input data or the number of cases, the coupling weights among the neurons in a large scale neural network can be exactly determined in a short time.

(2) By using the iterative learning together with the analytical approach, the significant reduction of the learning time is attained.

(3) Since the number of times of calculation is significantly reduced. The recognition or diagnosis may be implemented in an economical notebook type personal computer depending on a memory capacity thereof.

(4) Based on data of a particular form after the classification, the number of cases to be compared for the similarity is smaller after the classification and the comparison is facilitated and the precision of diagnosis is naturally improved.

(5) Since data in the appropriate range of scatter is accepted as a correct solution, the precision of diagnosis to unlearned data is naturally improved.

(6) When a result of diagnosis is prepared and if it is a correct solution, the diagnosis data and the correct solution are taken as the learning data for the next run. Accordingly, the precision of diagnosis is naturally improved gradually.

(7) By effecting the diagnosis logically more strict, the precision of diagnosis is further improved.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows a configuration of a conventional neural network;

Fig. 2 shows a configuration of the neural network in accordance with an embodiment of the present invention;

Fig. 3 shows a flow chart of an operation of a network configuring unit shown in Fig. 2;

Fig. 4 shows a relation between sums of input data in the respective cases and teacher data in which the teacher data changes once;

Fig. 5 shows a configuration of the neural network of the present invention when the teacher data changes once;

Fig. 6 shows a relation between total sums of input data in the respective cases and the teacher data in which the teacher data changes twice and a merging method therefor;

Fig. 7 shows a configuration of the neural network of the present invention when the teacher data changes twice;

Fig. 8 shows a relation between total sums of input data in the respective cases and the teacher data in which the teacher data changes three times and a merging method therefor;

Fig. 9 shows a configuration of the neural network of the present invention when the teacher data changes three times;

Fig. 10 shows other relation between total sums of input data in the respective cases and teacher data in which the teacher data changes once;

Fig. 11 shows other relation between total sums of input data in the respective cases and the teacher data in which the teacher data changes twice and a merging method therefor;

Fig. 12 shows other relation between total sums of input data in the respective cases and the teacher data in which the teacher data changes three times and a merging method therefor;  
 Fig. 13 shows conversion between values being operated and the input and output;  
 Figs. 14A to 14I show classification of learning data;  
 Fig. 15 shows learning data with consideration to a scatter;  
 Fig. 16 shows a case in which teacher data changes one time;  
 Fig. 17 shows a neural network in which teacher data changes one time;  
 Fig. 18 shows learning data which mostly overlap;  
 Fig. 19 shows learning data which mostly overlap;  
 Fig. 20 shows a method (No. 1) in which some data is used by reciprocals;  
 Fig. 21 shows a method (No. 2) in which some data is used by reciprocals;  
 Fig. 22 show data of four cases having scatter;  
 Fig. 23 shows a method (No. 1) of combination for teacher data which changed three times;  
 Fig. 24 shows a method (No. 2) of combination of teacher data which changes three times;  
 Fig. 25 shows a configuration of a neural network (in which teacher data changes three time);  
 Fig. 26 shows a method for taking diagnosis data and correct solution therefor into learning data;  
 Fig. 27 shows learning data and diagnosis data;  
 Fig. 28 shows learning data and diagnosis data with consideration to a scatter; and  
 Fig. 29 shows coupling weights to an input in determining a second output.

#### DESCRIPTION OF THE PREFERRED EMBODIMENTS

A diagnosis/control neural network system 1 of the present invention is now explained in detail with reference to the accompanying drawings.

First, a method for configuring the neural network of the present invention is now explained with reference to Fig. 3.

In a step S202, the sample data sets and teacher data set for the respective cases are inputted. The sample data of each set for the case  $X_i$  ( $i = 1$  to  $n$ ) is any value of  $0 \leq X_i \leq 1$ , and the teacher data for the case is either "0" or "1". It has been found out that coupling weights  $W_1, W_2, W_3, \dots, W_n$  may be assumed to be equal to each other as shown by the following equation (3) in most cases although there is a very rare exception (which will be discussed hereinafter).

$$W_1 = W_2 = W_3 \dots = W_n \quad (3)$$

A weight  $W_{n+1}$  for a threshold modification input "1" usually assumes another value.

Taking the equation (3) into account, in the equation (1), a sum  $\Sigma X_i$  of the input sample data  $X_i$  is first calculated and then a coupling weight may be multiplied thereto. In a step S204, a sum of the inputs is calculated for each data set corresponding to each case. In a step S206, they are rearranged in an order of the total sum  $\Sigma X_i$  and they are displayed together with the corresponding teacher data.

Table 1

Case Number	$\Sigma X_i$	Teacher Data
1'	$S_1$	0
2'	$S_2$	1
3'	$S_3$	.
.	.	.
n'	$S_n$	.

After the rearrangement, the number of neurons is determined in a step S208, and the coupling weights to the neurons are calculated in a step S210. Then, they are set in the neural network unit in a step S212. It has been found thorough various considerations that the treatment in determining the coupling weights should be changed depending on whether the teacher data to the maximum total sum (which is designated

by  $S_1$ ) is "0" or "1". First, it is assumed that the teacher data for  $S_1$  is "0".

For the convenience of explanation, a simple case will be discussed first.

In Fig. 4, an ordinate represents  $S_1, S_2, S_3, \dots, S_n$  of Table 1 and an abscissa represents "0" and "1" of the teacher data. In Fig. 4, the teacher data changes from "0" to "1" at a boundary of  $S_1$  and  $S_2$ . (It may change at other point). In the case of such a simple change, a hidden layer of the neural network is not  
 5 necessary at all and the network can be configured simply as shown in Fig. 5. Coupling weights  $W_a$  and  $W_b$  of the neural network are determined in the following manner.

Looking at the change of the teacher data in Table 1, the sums of the inputs immediately before and after the boundary at which the teacher data changes as shown in Table 1, that is,  $S_1$  and  $S_2$  in the present  
 10 example, are used to form the following linear simultaneous equations (4) and (5) with two unknowns.

$$S_1 \cdot W_a + 1 \cdot W_b = -10 \quad (4)$$

$$S_2 \cdot W_a + 1 \cdot W_b = 10 \quad (5)$$

15

The left terms represent inputs in the cases 1' and 2', and the right terms are the values of the inputs with which the output calculated by the formula (2) is practically "0" or "1" (Where  $U_1 = -10$ ,  $1/(1 + \exp(-10)) \approx 0$ , and where  $U_1 = 10$ ,  $1/(\exp(-10)) \approx 1$ ).

Assuming  $S_1 - S_2 \neq 0$ , the above simultaneous equations are solved.

20

$$W_a = -20/(S_1 - S_2) \quad (6)$$

$$W_b = 10 + 20S_2/(S_1 - S_2) \quad (7)$$

25

Since  $S_1 > S_2$ ,  $W_a$  is negative and  $W_b$  is positive. If  $S'$  is smaller than  $S_2$ , the following inequality is met.

$$S' \cdot W_a + 1 \cdot W_b > 10 \quad (8)$$

Thus, the outputs of  $S_3$  et seq in Fig. 4 are closer to "1" than  $S_2$  and they can be practically assumed  
 30 to be "1". It is seen that a set of simultaneous equations as shown by (4) and (5) are for one change in the teacher data.

As seen from Fig. 5, the coupling weights  $W_a$  and  $W_b$  are uniquely determined as the solutions of the simultaneous equations.  $\Sigma X_i$  is a variable but "1" is a fixed value. Where a sigmoid function is used as an output function of the neuron, a portion relating to "1" may be deleted from the circuit of Fig. 5 if a sigmoid  
 35 function having an offset of  $(1 \cdot W_b)$  along the abscissa is used.

Let's consider a case where the teacher data starts from "0", goes to "1" and returns to "0" again as shown by a segment 50 of Fig. 6. The teacher data changes two times. Since segments 51 and 52 can be generated in substantially same manner as that in Fig. 4, they are combined to generate the segment 50. Since the output is "1" only when both of the segments 51 and 52 are "1", it is seen that they may be  
 40 connected in an AND manner. The output of the AND circuit is "1" when both inputs are "1" so that a segment 53 is generated, which matches to the segment 50. A neural network for this case is shown in Fig. 7. Coupling weights of the input layer and the hidden layer are derived by solving simultaneous equations similar to the equations (4) and (5) at the two changes of the teacher data. In this case, consideration is made for the case where the teacher data changes from "0" to "1", and the case where it subsequently  
 45 changes from "1" to "0". The neuron shown by hatching at the bottom of the hidden layer is for threshold modification. When "20" is designated as the coupling weight between the input layer and the hidden layer as shown, the output is always practically "1". The coupling weights between the hidden layer and the output layer in an AND manner can be determined in a similar simple way. For example, they may be "20", "20" and "-30" as shown in Fig. 7.  $20 + 20 - 30 = +10$  and  $20 - 30 = -10$ , which correspond to "+1" and "0",  
 50 respectively. As seen, "-30" is a bias unit. An OR and (-10) shown in Fig. 7 will be explained hereinafter.

Let's consider a case where the teacher data changes three times as shown by a segment 70 of Fig. 8. Since segments 71 and 72 are generated in a similar manner to that described above, a segment 73 may be generated by ANDing them. A segment 74 is generated in a similar manner to that for the segment 71 and the segments 73 and 74 are coupled in an OR manner which outputs "1" when at least one of the  
 55 inputs thereto is "1". In this manner, a segment 75 is generated. It matches to the segment 70.

A neural network therefor is shown in Fig. 9. Again, the neuron shown by hatching at the bottom of the hidden layer is a bias unit and it always outputs "1". Third output from the top of the first hidden layer is directly connected to the output layer without routing the AND operation neuron. The coupling weights of

the OR operation neuron can be determined in a similar manner to that described above. They may be "20", "20" and "-10" as shown in Fig. 9 so that "1" is outputted when at least one of the inputs to the output layer is "1". (-10) is a bias unit. (OR) and (-10) of the hidden layer and (AND) and (-30) of the output layer shown in Fig. 9 will be explained hereinafter.

- 5 Similar analysis is conducted while the number of times of change of the teacher data is incremented to derive Table 2.

Table 2

10 The number of times of change of teacher data and the number of neurons of hidden layer

15	Number ( $n_1$ ) of times of change of teacher data	1	2	3	4	5	6	7	$n_1$
	Number ( $n_2$ ) of neurons of first hidden layer	(1)	2	3	4	5	6	7	$n_1$
20	Number ( $n_3$ ) of AND operation neurons of second hidden layer	0	(1)	1	2	2	3	3	$*_1$
25	Number ( $n_4$ ) of OR operation neurons of output layer	0	0	1	1	1	1	1	1

$$*_1: n_3 = \frac{n_1}{2} - \frac{1}{4} \{1 - (-1)^{n_1}\}$$

30 Table 2 shows a relation between the number of times of change of the teacher data and the number of neurons of the hidden layer. The number  $n_2$  of neurons of the first hidden layer, the number  $n_3$  of AND operation neurons (circuits) which are the second hidden layer, and the number  $n_4$  of OR operation neurons (circuits) which are the output layer are shown for the numbers "1" to "7" of times of change of the teacher data and a general number  $n_1$  (The neuron for modifying the threshold is separately provided for the hidden layer.). As shown in Table 2, the number of neurons in the first hidden layer is equal to the number  $n_1$  of times of change of the teacher data. Where the number  $n_1$  of times of change of the teacher data is "1", the neuron of the first hidden layer are the neuron of the output layer itself, and where the number  $n_1$  of times of change of the teacher data is "2", the neurons of the second hidden layer are the neurons of the output layer. It is convenient to consider that they are the neurons of the hidden layer for the number "1" in the parentheses of Table 2 than to express it as no hidden layer, when a method for calculating the coupling weight is to be considered. It has been proved by the analysis that it is sufficient to consider two hidden layers at most. As described above, when there are a plurality of outputs from the output layer, the above calculation method is repeated.

45 More generally, where the number of times of change of the teacher data is relatively large, the configuration of the neural network of the present invention is one shown in Fig. 2. Where the number of times of change of the teacher data is odd, there is a direct connection from the first hidden layer to the output layer without routing the second hidden layer. The coupling weights are determined in the manner described above.

50 Let's consider a case where the teacher data corresponding to the maximum sum  $S_1$  of the inputs is "1" as shown in Table 3.



Table 3

Case Number	$\Sigma X_i$	Teacher Data
1'	$S_1$	1
2'	$S_2$	0
3'	$S_3$	.
.	.	.
n'	$S_n$	.

Fig. 10 shows a case where the teacher data changes once. In Fig. 10, the teacher data changes between  $S_1$  and  $S_2$  although it may change at other point. The neural network in this case requires no hidden layer and it may be the same one as that shown in Fig. 5 but the coupling weights are different, which are determined by simultaneous equations similar to the equations (4) and (5) but the sign of the right term is changed to render the outputs to be "1" and "0".

Fig. 11 shows a case where the teacher data changes twice as shown by a segment 100. Since segments 101 and 102 are generated in the same manner as that described above, the segment 100 may be generated based thereon. When at least one of the segments 101 and 102 is "1", the output is to be "1". That is, they are coupled in an OR manner. In this case, the neural network is substantially identical to that of Fig. 7 but the output layer functions as the OR operation instead of the AND operation and the coupling weight for the threshold modification is changed from "-30" to "-10", for example.

Fig. 12 shows a case where the teacher data changes three times as shown by a segment 110. In this case, segments 111 and 112 are generated in the same manner as that described above, and they are ORed to generate a segment 113. Then, a segment 114 is generated in the same manner as that for the segment 111. Finally, the segment 113 and the segment 114 are coupled in an AND manner to generate a segment 115. It matches to the segment 110. In this case, the neural network is substantially identical to that of Fig. 9 but the OR operation neuron is arranged in the second hidden layer and the coupling weight for modifying the threshold is changed so that the output layer functions as the AND operation. The result of the analysis while incrementing the number of time of change of the teacher data is summarized in Table 4.

Table 4

Number ( $n_5$ ) of times of change of teacher data	1	2	3	4	5	6	7	$n_5$
Number ( $n_6$ ) of neurons in first hidden layer	(1)	2	3	4	5	6	7	$n_5$
Number ( $n_7$ ) if OR operation neurons of second hidden layer	0	(1)	1	2	2	3	3	$*_2$
Number ( $n_8$ ) of AND operation neurons of output layer	0	0	1	1	1	1	1	1

$$*_2: n_7 = \frac{n_5}{2} - \frac{1}{4} \{1 - (-1)^{n_5}\}$$

Table 4 shows the number  $n_6$  of the neurons of the first hidden layer, the number  $n_7$  of OR operation neurons as the number of neurons of the second hidden layer, and the number  $n_8$  of the AND operation neurons in the output layer, for the numbers "1" to "7" of times of change of the teacher data and a

general number  $n_5$  when the teacher data of  $S_1$  is "1". Like in Table 2, the neuron for modifying the threshold is separately provided for the hidden layer and the number "1" in the parentheses also represents the outputs of the output layer. By comparing Table 2 and Table 4, it is seen that the AND and OR are simply replaced. In this case, two hidden layers at most may be considered.

5 More generally, where the number of times of change of the teacher data is relatively large, the configuration of the neural network is like that shown in Fig. 2. When the number of times of change of the teacher data is odd, there is a case where direct connection is made from the first hidden layer to the output layer without routing the second hidden layer as shown by a broken line. The second hidden layer includes the OR operation neurons and the output layer includes the AND operation neurons. The method  
10 for calculating the coupling weights is same as that described above.

Where a plurality of outputs are required for the output layer, the analysis is repeated for each output in the process described above. When the outputs are to be determined, a portion of the neural network may be shared if the boundary of the change of the teacher data shown in Table 1 and Table 2 are same. Thus, the entire configuration may be simplified by utilizing this fact.

15 In the equations (6) and (7), when  $S_1$  and  $S_2$  are equal, the denominators of those equations are zero. In such a case, instead of simply summing the inputs, a number  $K_i$  which is determined from a random number and which is not "1" is multiplied to each input before the summation. Namely,  $\sum K_i X_i$  are calculated and then they are handled in the same manner as  $S_1, S_2, S_3, \dots, S_n$  of Table 1 and Table 3. By doing so, it rarely occurs that the denominators of the equations (6) and (7) are zero, except when the inputs of the  
20 plurality of cases are totally equal. If the denominators of the equations (6) and (7) are still zero, a process of deriving another random number may be repeated. As is apparent from the previous description, the fact that the denominator is zero causes a problem only when  $S_1$  and  $S_2$  are equal and the corresponding teacher data are different. Even if  $S_1$  and  $S_2$  are equal, no problem is raised if the corresponding teacher data are equal, since they are not used for the calculation of the coupling weights. Accordingly, more  
25 generally, in table 2 and Table 4,  $\sum X_i$  is not calculated as a simple summation but  $\sum K_i X_i$  is calculated by using the number  $K_i$  determined from the random number and the calculation may be proceeded in the manner described above. In this case, the coupling weight for each input  $X_i$  may be considered to be  $K_i X_i$ . When  $K_i$  is used as the weight factor, it can be readily attained by adding a neuron to the input layer. It is assumed that  $S_1'$  and  $S_3'$  are equal, the corresponding teacher data are equal, and there is a sum  $S_2'$  which is between  $S_1'$  and  $S_3'$  and has the teacher data equal to the teacher data for  $S_1'$  and  $S_3'$ . In such a  
30 case, as is apparent from the above equations, there is no change among the three teacher data and hence  $S_2'$  is not used for the calculation of the coupling weight. Accordingly, it is seen that the data  $S_2'$  does not exist as the diagnosis data and it is not worthwhile adding. It may be deleted to save the memory space. If the newly added teacher data changes the boundary of the change of the teacher data, it is worthwhile  
35 adding.

In accordance with the present method, the following advantages are offered:

- (1) The iterative learning required in the prior art is not necessary.
- (2) Since the number of hidden layers, the number of neurons required therefor and the coupling weights can be analytically determined, highly efficient diagnosis is attained in a short time.
- 40 (3) It has been proved that the number of hidden layers may be two at maximum.
- (4) Since the coupling weights are analytically determined in a simple manner, it is applicable to a large scale system having a number of input data and output data.
- (5) Since the iterative learning required in the prior art is not necessary, more economical computer may be used for the recognition and diagnosis.
- 45 (6) When the input data is to be added to enhance the precision of diagnosis, it can be determined whether it is worthwhile adding or not based on the input data and the teacher data and if it is not worth, the addition is stopped. Accordingly, the memory can be efficiently used.

In the equations (4) and (5), the output "0" or "1" is approximated from the neuron. In a conversion shown in Fig. 14, considering  $\epsilon$ , a positive number close to zero,  $\epsilon$  and  $(1-\epsilon)$  are used for the intermediate  
50 calculation instead of "0" and "1", and in a final output stage, it is returned to a number in the range from "0" to "1". When this is applied to the right terms of the equations (4) and (5), they are expressed by

$$S_1 \cdot W_a + 1 \cdot W_b = \log(\epsilon/(1-\epsilon)) \quad (9)$$

$$55 \quad S_2 \cdot W_a + 1 \cdot W_b = -\log(\epsilon/(1-\epsilon)) \quad (10)$$

where log is natural logarithm. In the method in which the output of the equation (2) is treated as "0" or "1" for practical reason, approximation calculation is partially included and hence an error is introduced. The

present method includes no such error.

A further embodiment of the present invention is explained with reference to Figs. 14A to 14I which illustrate the classification of the learning data. An abscissa represents an input data number  $n$  of the same case and an ordinate represents a magnitude of corresponding input data  $x$ . A number of examples are classified based on characteristics thereof as shown in Figs. 14A to 14I. In Figs. 14A to 14C, there is one peak, and they may be classified in detail based on the magnitude of peak, the value  $n$  for the peak and a partial gradient (convex upwardly or concave downwardly) although it is not shown. In Figs. 14D to 14F, there are two peaks, and they may be classified based on the magnitudes of the left and right peaks, the magnitudes of the peaks and valleys, and the value  $n$  for the peak or valley (whether the  $n$  for the valley is closer to the  $n$  for the left peak or the  $n$  for the right peak) although it is not shown. In Figs. 14G to 14I, there are three peaks and they may be classified from various aspects as described above. The classification may be also made based on shapes such as rectangular and pedestal although it is not shown.

The entire input data for learning is classified in accordance with Figs. 14A to 14I, and a configuration of a multi-layer neural network and coupling weights among the neurons may be analytically determined in accordance with the above method. In the above analytical method, the configuration of the neural network and the coupling weights among the neurons are normally determined in several to several tens seconds although it depends on the scale of data. Since the number of cases belonging to a particular class is smaller than that when no classification is made, the comparison is facilitated in the diagnosis and the discrimination of confusing and similar data is relatively easy, and the precision of diagnosis is improved. When the iterative learning method which requires a number of times of learning by the prior art back propagation learning method is learnt after such classification, the time required to learn about all types may increase in many times. Accordingly, it is generally difficult to apply the present method to back propagation learning method after the classification of the input data. However, depending on the input data, it may be utilized for the back propagation learning method to reduce the overall learning time.

A further embodiment of the present invention is explained with reference to Figs. 15 and 16. In the present embodiment, the analytical method explained in Figs. 2 to 13 is expanded with the consideration of the scatter of the input data.

In Fig. 15, an abscissa represents an input data number  $n$  and an ordinate represents a value  $x$  of the input, as they do in Figs. 14A to 14I. Two measured curves  $F$  and  $G$  with scatter are shown. An upper limit of  $F$  is represented by  $F_u$ , a lower limit by  $F_d$ , an upper limit of  $G$  by  $G_u$  and a lower limit by  $G_d$ .

In the explanation of the present invention, it is assumed that the input  $x$  is any positive value, the number of output (teacher data) of the neural network is 1, and the value thereof is "0" or "1". Where there are more than one output, each output is handled as follows.

It is assumed that when  $F$  is inputted, the neural network outputs "0", and when  $G$  is inputted, it outputs "1".

Total sums of inputs  $Sf_u$ ,  $Sf_d$ ,  $Sg_u$  and  $Sg_d$  excluding the threshold modifying input with arbitrarily determined weights (coupling weights)  $K_i$  are determined by formula (11) as they

$$\left. \begin{aligned} Sf_u &= \sum_{i=1}^n F_{ui} \cdot K_i \\ Sf_d &= \sum_{i=1}^n F_{di} \cdot K_i \\ Sg_u &= \sum_{i=1}^n G_{ui} \cdot K_i \\ Sg_d &= \sum_{i=1}^n G_{di} \cdot K_i \end{aligned} \right\} \quad (11)$$

are in Table 1.

Table 5

Input Total Sums and Teacher Data Rearranged in the Order of Magnitude			
Case No.	Total Input Sum	Teacher Data	Remark
1'	$Sf_u$	0	Substitute by $S_1$
2'	$Sf_d$	0	Substitute by $S_2$
3'	$Sg_u$	1	Substitute by $S_3$
4'	$Sg_d$	1	Substitute by $S_4$

As shown in Table 5, the data are rearranged in the order of magnitude and they are displayed together with the corresponding teacher data. After such rearrangement, it is necessary to change the weights (coupling weights) of the second hidden layer et seq of the neural network depending on whether the teacher data having the maximum input total sum is "0" or "1". For the input F and G having a scatter, in order to get a correct solution as an output of the neural network if it is within the range of scatter, it is necessary that either a formula (12) or a formula (13) is met when the weight (coupling weight) to the input is positive or "0".

$$Sf_u \geq Sf_d \geq Sg_u \geq Sg_d \quad (12)$$

$$Sg_u \geq Sg_d \geq Sf_u \geq Sf_d \quad (13)$$

Depending on the manner of change of a curve, the formula (12) or (13) may be met even if all weights (coupling weights) to the inputs x are selected equal. However, depending on the curve, the value of other curve (for example,  $Sg_u$ ) may fall between sums of one curve (for example,  $Sf_u$  and  $Sf_d$ ). In order to assure that the magnitude relation of the formula (12) or (13) is positively met, the following methodology is used. First, as candidates for the weights (coupling weights)  $K_i$  to be used, 0.0001, 0.001, 0.01, 0.1, 1.0, 10.0, 100.0, 1000.0, 10000.0 etc. are prepared, and the right side of the crosspoint of the curves F and G of Fig. 15 is weighted with very large weights (coupling weights)  $K_i$  compared to the left side (for example, 10 to 100 times) so that the total sums of inputs multiplied by the weights (coupling weights) are emphasized on the right side of the crosspoint of the curves F and G and the magnitude relation of the formula (12) in which the sum for the curve F is larger is met.

On the other hand, if the left side of the crosspoint of the curves of Fig. 15 is weighted with very large weights (coupling weights)  $K_i$  compared to the right side (for example, 10 to 100 times), the total sums of the inputs multiplied by the weights (coupling weights) are emphasized on the left side of the crosspoint of the curves F and G so that the magnitude relation of the formula (13) in which the sum of the curve G is larger is met.

A magnitude relation should not be such that a total sum determined by other curve falls between two total sums determined by one curve.

When the weight (coupling weight) is positive or "0" for the data in the range of scatter shown for the curve F, the sum of the products of the inputs x and the weights (coupling weights)  $K_i$  is always between  $Sf_u$  and  $Sd_d$ . Similarly, when the weight (coupling weight) is positive or "0" for the data in the range of scatter shown for the curve G, the sum of the products of the inputs x and the weights (coupling weights)  $K_i$  is always between  $Sg_u$  and  $Sg_d$ . Accordingly, as will be explained later, by selecting the weights (coupling weights) to meet the formula (12) or (13), the same output as that designated by the curves  $F_u$  and  $F_d$  is attained for the input x in the range of scatter of the curve F. Similarly, the same output as that designated by the curves  $G_u$  and  $G_d$  is attained for the input x in the range of scatter of the curve G.

For the purpose of simplification, it is assumed that the formula (12) is met, the output (teacher data) is "0" for the curve F and the output (teacher data) is "1" for the curve G.

In order to make it easy to process the formulas of the Table 5, the characters are substituted as follows.

$$Sf_u \rightarrow S_1, Sf_d \rightarrow S_2, Sg_u \rightarrow S_3, Sg_d \rightarrow S_4$$

In Fig. 16, an ordinate represents  $S_1, S_2, S_3$  and  $S_4$  of Table 5 and an abscissa represents the teacher data. The teacher data changes only once at the boundary of  $S_2$  and  $S_3$ . In order to generate this simple

change, no hidden layer is required in the neural network and a simple construction shown in Fig. 17 is sufficient. This is different from Fig. 5 in that the inputs are weighted. The coupling weights  $W_a$  and  $W_b$  of the neural network are determined as follows.

- Based on the change of the teacher data of Table 5, the following simultaneous equations are prepared by using the input total sums  $S_2$  and  $S_3$  immediately before and after the boundary of change of the teacher data shown by broken lines in the table while referencing to Fig. 17.

$$S_2 \cdot W_a + 1 \cdot W_b = -10 \quad (14)$$

$$S_3 \cdot W_a + 1 \cdot W_b = 10 \quad (15)$$

The value of the left term is the input to the output layer, and the value of the right term is an input value which allows that the output calculated by the formulas (1) and (2) be practically considered as "0" or "1".

- (Because  $1/(1 + \exp(-(-10))) \approx 0$ ,  $1/(1 + \exp(-(10))) \approx 1$ , see formulas (1) and (2))  
By solving the above simultaneous equations ( $S_2 - S_3 \neq 0$ ),

$$W_a = -20/(S_2 - S_3) \quad (16)$$

$$W_b = 10 + 20S_3/(S_2 - S_3) \quad (17)$$

Since  $W_a$  is negative and  $W_b$  is positive, the following inequality is met if  $S'$  is smaller than  $S_3$ .

$$S' \cdot W_a + 1 \cdot W_b > 10 \quad (18)$$

Accordingly, the output to the input total sum which is smaller than  $S_3$  of Fig. 16 is closer to "1" than that for  $S_3$  and it may be practically considered as "1".

When  $S''$  is larger than  $S_2$ , the following inequality is met.

$$S'' \cdot W_a + 1 \cdot W_b < -10 \quad (19)$$

Accordingly, the output to the total input sum which is larger than  $S_2$  of Fig. 17 is closer to "0" than that for  $S_2$  and it may be practically considered as "0".

Thus, for the input in the range of scatter of the curve F or G of Fig. 15, "0" or "1" output is produced.

- Fig. 18 shows an example in which the curves F and G with scatter overlap in most areas. In such a case, if the weights (coupling weights)  $K_i$  for the inputs are equal, the total sums of the inputs weighted by the coupling weights are almost equal depending on the shapes of the curves and the magnitude relation of the input total sums may not meet the above required relation.

- However, if the weight (coupling weight)  $K_i$  at one point in the non-overlapping area of the curves F and G is set sufficiently larger than the weights (coupling weights)  $K_i$  of other area (for example, 10 to 100 times), the affect of the value of that area appears prominent so that the magnitude relation of the formula (12) or (13) is met. For example, if only the weight (coupling weight)  $K_i$  in the range A of Fig. 18 is selected sufficiently large, the sum for the curve H becomes dominant and the formula (12) is met, and if only the coupling weight in the range B is selected sufficiently large, the sum for the curve G becomes dominant so that the formula (13) is met. If the weight  $K_i$  in the range A is selected larger than those of other area and the weight  $K_i$  in the range B is selected smaller than those of other area, the curve F is more dominant. The opposite selection may be made.

- Fig. 19 shows another approach for a similar case. In order to simplify the expression, the range of scatter is omitted in the drawing. In Fig. 19, again the abscissa represents the data number  $n$  and the ordinate represents the input data  $x$ . Two type of data are shown, one by circle and the other by X. Most data are overlapped but three data are not coincide to each other. Two data marked with X are smaller than the data marked with circle and one data marked with X is larger than the data marked with circle. In this case, if the scatter not shown is considered, the total sums of the products of the input data  $x$  and the weights (coupling factors)  $K_i$  of the respective cases are almost equal if the weights (coupling factors)  $K_i$  are simply made equal and the magnitude relation of the formula (12) or (13) may not be met. The approach to partially increase the coupling weights to the input data has been described above. In Figs. 19-21, another method for meeting the magnitude relation of the formula (12) or (13) even if all coupling weights to the input data are equal is shown.

In Fig. 20, the data marked with X which is larger than the data marked with circle is not used as it is but the data marked with circle and X for that data number  $n$  are used by reciprocals. The reciprocal of the data marked with circle is represented by dotted circle. The reciprocal of the data marked with X is represented by  $\Delta$ . In the case classified by the mark X, two data marked with X and one data marked with  $\Delta$  are smaller than the data marked with circle (including dotted circle) of the same data number. Accordingly, even if all weights (coupling weights)  $K_i$  are selected equal, the total sum of the input marked with circle is larger and it can be discriminated.

In Fig. 21, the only one data marked with X which is larger than the data marked with circle is left as it is, and the two inputs for the  $n$  marked with X which are smaller than the data marked with circle as well as the data marked with circle are used by reciprocals. As a result, two data marked with  $\Delta$  and one data marked with X are larger than the data marked with circle of the same data number. Accordingly, even if all weights (coupling weights)  $K_i$  are selected equal, the total sum of the inputs marked with circle is smaller and it can be discriminated.

Fig. 22 shows four cases having scatters. In this case, a range of the data numbers such as a range A in which the data of the four cases assume different values from each other is selected, and the weights (coupling weights)  $K_i$  for the input data in that range are selected sufficiently larger than those of other range. Thus, the total sums of the inputs weighted by the weights (coupling weights)  $K_i$  are larger in the order of F, G, H and I, and the sum of the upper limits of the scatter of the data is smaller than the lower limit and the order of the total sums of the four cases is not disordered.

Assuming that the output (teacher) data for the input data F, G, H and I are "0", "1", "0" and "1", the relation of the input total sums and the teacher data in the order of magnitude of the input total sum is shown in Table 6. The symbols for the input total sums are shown by  $Sf_u$ ,  $Sf_d$ ,  $Sg_u$ ,  $Sg_d$ ,  $Sh_u$ ,  $Sh_d$ ,  $Si_u$ ,  $Si_d$  with the suffixes  $u$  and  $d$  representing the upper limit and the power limit, respectively, as they do in the previous example. This relation may be represented by a segment graph as shown in Fig. 23.

As shown by a segment 220, the teacher data starts from 0, passes through 1 and returns again to 0 and finally becomes 1. the teacher data changes three times in this example. Segments 221 and 222 are generated in substantially same manner as that of Fig. 4 and they are combined to generate a segment 223. Since the output may be "1" only when both of the segments 221 and 222 are "1", it is seen that they may be connected to an AND circuit. A segment 224 is generated in the same manner as that for the segment 221. Finally, the segments 223 and 224 are combined by an OR circuit. Since the output of the OR circuit is "1" when at least one of the inputs is "1", a segment 225 is generated. It coincides with the segment 220.

If the output (teacher) data for the input data F, G, H and I are reversed to "1", "0", "1" and "0", respectively, the relation between the input total sum and the teacher data is that shown in Table 7. This relation may be represented by a segment graph shown in Fig. 24. As shown by a segment 230, the teacher data starts from "1", passes through "0", returns again to "1" and finally becomes "0". The teacher data changes three times in this example. Since segments 231 and 232 are generated in the same manner as that of Fig. 4, the segments 231 and 232 are connected to an OR circuit to synthesize a segment 233. Then, a segment 234 is generated in the same manner as that for the segment 231. Finally, the segments 233 and 234 are combined in an AND circuit to generate a segment 235. It coincides with the segment 230.

In the above example, if the total sum of the products of the inputs and the weights (coupling weights)  $K_i$  is between  $Sf_u$  and  $Sf_d$ , between  $Sg_u$  and  $Sg_d$ , between  $Sh_u$  and  $Sh_d$ , or between  $Si_u$  and  $Si_d$ , the output (teacher) data is designated. However, the output data is not designated when the total sum is between different curves, that is, between  $Sf_d$  and  $Sg_u$ , between  $Sg_d$  and  $Sh_u$ , or between  $Sh_d$  and  $Si_u$ . In this case, the following process may be effected.

(a) Values of the both boundaries are outputted in accordance with a given formula.

(b) "?" is added in the output display to indicate non-inference.

As an example of (a), a value determined linearly or by a sigmoid function between "0" and "1" is outputted for a non-designated range between the output designations "0" and "1" in accordance with a deviation from the boundary.

Fig. 25 shows a configuration of the neural network corresponding to Figs. 23 and 24. In a bottom line of the hidden layer, a neuron shown by hatching which always outputs "1" for modifying a threshold is coupled, and the input of the first hidden layer is coupled to only the threshold modifying input "1".

When the total sum of the inputs weighted by the weights (coupling factors)  $K_i$  corresponds to the teacher data "0" as shown in Fig. 23, it is required that the second hidden layer functions as the AND circuit and the output layer functions as the OR circuit. On the other hand, when the maximum total sum if the inputs weighted by the weights (coupling weights)  $K_i$  corresponds to the teacher data "1" as shown in Fig. 24, it is required that the second hidden layer functions as the OR circuit and the output layer functions

as the AND circuit.

As seen from the those figures, only the coupling weight for modifying the threshold changes between the AND circuit and the OR circuit. An equivalent coupling weight to the input  $x$  is a product of the weight (coupling weight)  $K_i$  used in the formula (11) and the coupling weights  $W_a$  and  $W_b$  determined by the formulas (16) and (17).

Table 6

Four input cases in which teacher data starts from 0		
Case No.	Input Total Sum	Teacher Data
1'	$Sf_u$ $Sf_d$	0 0
2'	$Sg_u$ $Sg_d$	1 1
3'	$Sh_u$ $Sh_d$	0 0
4'	$Si_u$ $Si_d$	1 1

Table 7

Four input cases in which teacher data starts from 1		
Case No.	Input Total Sum	Teacher Data
1'	$Sf_u$ $Sf_d$	1 1
2'	$Sg_u$ $Sg_d$	0 0
3'	$Sh_u$ $Sh_d$	1 1
4'	$Si_u$ $Si_d$	0 0

The present invention is also applicable to cases similar to Tables 6 and 7 where the teacher data changes in more complex and more frequent manner.

Another embodiment which is applicable when a correct solution is known as the output for the diagnosis data is now explained with reference to a flow chart of Fig. 26.

When the calculation is started, learning data, teacher data therefor and a number of values which are candidates for the weights (coupling weights)  $K_i$  are read as learning data (1) from a memory, a hard disk (HD) or a flexible disk (FD) of a computer.

Then, whether data to be added as the leaning data (2) is present or not is determined. Since there is no additional data initially (TUIKAS=NASHIS), the calculation of a total sum is initiated. Total sums of inputs weighted by appropriate weights (coupling factors)  $K_i$  in the manner described above are determined and they are arranged together with the corresponding teacher data in the order of magnitude of the input total sum, and the number of times of change of the teacher data is considered. An output layer comprising the number of neurons of the first and second hidden layers determined by the number of times of change of the teacher data plus one neuron is prepared and the coupling weights are determined in the manner described above. In this step, the step corresponding to the learning in the prior art reverse propagation learning method is completed.

Then, a screen of a CRT (cathode ray tube) inquires the need for diagnosis. If Y (YES) is entered and the diagnosis data is inputted, the corresponding values are read from the memory, hard disk (HD) or

flexible disk (FD) of the computer. Then, they are diagnosed and the result is displayed. Then, if Y is entered to the inquiry "Correct solution known?" on the CRT, the characters TUIKAS are changed to ARIS and the correct solution as well as the data inputted for the diagnosis are saved in the HD or FD as the learning data (2). Then, if Y is entered to the inquiry "Terminate?" on the CRT screen, the diagnosis is terminated.

While not shown, when the configuration of the neural network and the coupling weights have been determined, various values including the coupling weights may be saved in the HD or FD and necessary calculation loop may be modified for use in the next diagnosis to improve the efficiency.

When the calculation is resumed in the next run, the calculation is executed in the substantially same manner. As to the data diagnosed in the previous run and for which the correct solution has been inputted, it is added to the input data in the step of the learning data (2) so that the learning is done better and the correct solution is naturally given to the same inquiry as that of the diagnosis data correctly solved in the previous run. In this manner, the precision of diagnosis is enhanced gradually. In the present system in which the configuration of the neural network and the coupling weights are analytically determined, the diagnosis may be resumed with a higher precision in several to several tens seconds after the addition of the learning data, although it depends on the scale of the number of data.

Table 8 show various propositions used in general inference.

Table 8

- (1) If A, then B
- (2) If not B, then not A
- (3) If B, then A

For the proposition (1) of Table 8, the proposition (2) of Table 8 is referred to as a contraposition which is a proposition which uses the negation of a conclusion for one proposition as a hypothesis and the negation of the hypothesis as a conclusion. The true and false of the original proposition and the contraposition always match. On the other hand, the proposition (3) of Table 8 is referred to as a converse to (1). The converse is a proposition derived by inverting the hypothesis and the conclusion of one proposition. Even if the original proposition is true, the converse is not always true.

In the prior art neural network, which of the input learning data the diagnosis data is similar to is determined depending on which of the previously learnt output (teacher) data the output data at the time of input of the diagnosis data is similar to. As seen from the general description of Table 8, when (1) of Table 8 is "If the learning data is A, then the output data is B", the use of (2) of Table 8 is always true but the use of (3) of Table 8 is not always true, in a strict sense. (3) of table 8 includes those whose conclusions are A if the hypothesis is B and those whose conclusions are not B. However, in the neural network, there is a case in which a conclusion which is not A but similar to A may be treated as A without problem.

An example is given. Fig. 27 shows input data of a neural network in one case. An abscissa represents a date number n and an ordinate represents input data x. The learning of data marked with circle in Fig. 27 corresponds to A of Table 8 (1) and the output (teacher) data therefor corresponds to B of Table 8 (1). It is assumed that the diagnosis data are mostly equal to the data marked with circle but only data numbers i and k are different. In Fig. 27, the data which coincide with the data marked with circle are omitted for the expression of mark X. In the diagnosis data, the i-th and k-th data are represented by  $(x_i + \Delta x_i)$  and  $(x_k + \Delta x_k)$ , and the weights (coupling weights) for the input data are represented by  $K_i$  and  $K_k$ . A difference  $\Delta S$  between total sums of the two inputs weighted by the weights (coupling weights) is given by:

$$\Delta S = \Delta x_i \cdot K_i + \Delta x_k \cdot K_k \quad (20)$$

If  $K_i$  and  $K_k$  have been set to appropriate values for the learning data marked with circle of Fig. 27, it is easy to give one of  $\Delta x_i$  and  $\Delta x_k$  to determine the other so that  $\Delta S$  becomes zero. In this case, even if the learning has been made by using only the learning data marked with circle and even if the data which is partially out of the range of the circle marks of Fig. 27 is diagnosed, the total sums weighted by the weights (coupling weights) are equal for the case of learning and the case of diagnosis, and hence the output data are equal for both cases. Namely, even if the output data are equal, the input data are not always equal. This correspond to "the converse is not always true" as described in conjunction with Table 8. In this case, some are inferred from the output data as anticipated and the other are non-anticipated ones. However, in the neural network, even if unlearned data is inputted, it may be regarded to be almost same as the learnt data if the output is very similar to that for the learnt data. In order to further enhance the precision of



diagnosis, however, it is necessary to distinguish it from the output resulting in nonanticipated input data. A method for positively eliminating the nonanticipated data is explained with reference to Fig. 27.

The data for the diagnosis is inputted to calculate the output and most probable learning data is inferred from a number of learnt cases. It is assumed that the data inferred to be most probable is the data marked with circle in Fig. 27. Then, whether the data marked with circle in Fig. 27 match to the diagnosis data or not is determined for each data number. If the learning data and the diagnosis data match each other for all of the input data, it is determined that the diagnosis data is equal to one for the selected learning case. If a portion of the input data does not match, it is determined that the diagnosis data is different from one for the selected learning case. In this manner, the precision of diagnosis is enhanced.

When it is determined by the above diagnosis that the diagnosis data and the learning data do not completely match, "No match with learning data" may be indicated as well as "Similar cases" with a percentage of the number of matching data to the total number of inputs per case may be indicated. This will give more kind reply to a user than simply indicating "No match with learning data".

In the above example, only two data do not match to the data marked with circle of Fig. 27. When more data do not match to the data marked with circle, a similar method may be applied.

The confirmation of the diagnosis when the scatter of the learning data is taken into consideration is now explained with reference to Fig. 28. In Fig. 28, an abscissa represents the data number and an ordinate represents the input data. A curve  $F_u$  shows upper limit input data of a curve  $F$  in one case, and a curve  $F_d$  shows lower limit data of the same case. It is assumed that the curves  $F_u$  and  $F_d$  of Fig. 28 have been selected as the most probable case after the input of the diagnosis data and the calculation of the output. Like in the case of Fig. 27, the following calculation for confirmation is made. Where  $n_1$  is the number of input data per case;

- (1) determine the number  $n_2$  of diagnosis data (marked with X) which are on or below the curve  $F_u$ , and
- (2) determine the number  $n_3$  of diagnosis data which are on or above the curve  $F_d$ .

If  $n_1$ ,  $n_2$  and  $n_3$  are all equal, it is determined that the diagnosis data are in the range of the curves  $F_u$  and  $F_d$  and it corresponds to the case of the curve  $F$ .

If at least one of  $n_2$  and  $n_3$  is not equal to  $n_1$ , it is determined that a portion of the diagnosis data is out of the range of scatter encircled by the curves  $F_u$  and  $F_d$ . In such a case, it is indicated that the diagnosis data do not completely match to the learning data as well as the percentage of the diagnosis data to the input data which are between the curves  $F_u$  and  $F_d$  are indicated as explained in Fig. 27 so that the degree of similarity to the learning data is indicated for the convenience of the user. In this manner, exact diagnosis may be attained even for the input data having the scatter.

A specific method for comparing the diagnosis data and the learning data may be the following one instead of the steps (a) and (b) described above.

- (a') Determine the number of diagnosis data which are above the curve  $F_u$  of Fig. 28.
- (b') Determine the number of diagnosis data which are below the curve  $F_d$  of Fig. 28.

If both of the numbers of data determined in the steps (a') and (b') are zero, it is determined that the diagnosis data is within the range of scatter of the learnt data for a case.

In Fig. 28, the upper limit  $F_u$  and the lower limit  $F_d$  of the range of scatter may be set somewhat larger than an actual scatter which may be experimentally determined in order to prepare for any future possible expansion of the scatter by some reason. If the diagnosis data does not completely match to the learning data, not only the number of data which are out of the range of scatter but also the degree of deviation may be included to affect to the indication.

The methods described in Figs. 27 and 28 may be applied to the following cases.

- (1) Application to a reverse learning method of a multi-layer neural network
- (2) Application to the present invention to analytically determine the configuration of the multilayer neural network and the coupling weights.
- (3) Not applied to the neural network but repeat the confirmation as described above as to the similarity of the diagnosis data to the learning data of all cases.

In any of the above cases, in order to compare the diagnosis data with the learning data, the respective numbers may be compared by prepared software, or a diagnosis data group and a learning data group of a case may be collectively compared by a matrix or a matrix formula to indicate a group having a value equal to or smaller than a value of other group, or count the number thereof.

The above cases (1) to (3) are explained in the reverse order of (3) to (1) for the convenience of explanation. In the method (3), to which learning case the diagnosis data is most similar (including totally match case) is determined for the input data of all cases. Accordingly, where the number of cases is large and the number of data per case is large, some time is required for the calculation for the diagnosis. However, in the prior art reverse propagation learning method, where the learning takes too much time or

the learning does not proceed because of local minimum, this method may be practiced if a high speed operation is not absolutely required. Since the correctness of diagnosis rather than the calculation speed is frequently required in the fault diagnosis, this method may be practical in many cases.

5 In the method (2), the diagnosis data is inputted, the output is calculated, the similar learning case is selected and the confirmation calculation is made in the manner described above. Since the number of times of calculation is smaller than that of the method (2), the calculation time is shorter and the diagnosis is more correct.

10 In the method (1), the confirmation calculation shown in Figs. 27 and 28 is required in the worst case (where the configuration of the neural network is very simple, for example, the number of hidden layer is one and the number of neuron of the hidden layer is one). However, where the neural network is more complex, the confirmation calculation is possibly not required. This is explained below assuming that the number of hidden layer is one.

Where the number of neurons of the first hidden layer is more than one, the coupling weights between the neurons of the input layer and the first neuron of the first hidden layer are generally different from the coupling weights between the input layer and the second neuron of the first hidden layer. Accordingly, even if  $\Delta S$  given by the formula (20) is zero with respect to the coupling weights between the input layer and the first neuron of the first hidden layer,  $\Delta S'$  calculated in the same manner as that for  $\Delta S$  with respect to the coupling weights between the input layer and the first hidden layer is highly possibly not zero. As a result, the output from the second neuron of the first hidden layer is highly possibly different between the learning data and the diagnosis data. This affect generally also appears in the output of the output layer. As a result, the learning data marked with circle shown in Fig. 27 and the diagnosis data (data which coincide with the circles are not shown) are highly probably discriminated. Where the configuration of the neural network is further complex, this trend is more prominent and the confirmation calculation described in connection with Figs. 27 and 28 is not required in many cases.

25 In the analytical method of the present invention, it is naturally desired to eliminate the confirmation calculation in the diagnosis as much as possible. This may be attained to some extent where there are a plurality of outputs. This is explained in connection with the network of Fig. 25 and a network of Fig. 29 having a similar configuration. As to the first output, the configuration of the neural network and the coupling weights are determined by the method explained in connection with Fig. 25. Initial weights (coupling weights) for the input are shown by  $K_i$ . When the second output of Fig. 29 is to be determined, the initial weights (coupling weights) for the input are set differently from those for the first output within the range which meets the respective requirements explained in connection with Fig. 25, and they are shown by  $K_i'$ . Where other outputs are present, the initial coupling weights for the input are set differently from those of the first and second output calculations in a similar manner. Thus, even if a difference between the learning data and the diagnosis data does not appear in the first output shown in Fig. 25, a difference may highly probably appear in the second output shown in Fig. 29 because the weights (coupling weights) to the input are changed. As a result, as a whole output data, the possibility of detection of difference between the learning data and the diagnosis data explained in Fig. 27 is increased.

40 In accordance with the present method, since the number of hidden layers and the number of neurons are known and the number of coupling weights to be learnt is significantly decreased, the leaning time can be reduced. Where a program of the reverse propagation learning is already present, the modification of the program is easy and various functions of the existing program can be utilized while the advantages of the present invention is reserved.

45 By the methods described above, the equivalent status to that attained by the conventional iterative learning is attained in an efficient manner. Accordingly, by using the input data used for the calculation, the calculation output which is totally equal to or very close to the teacher data can be outputted. If the number of cases which are handled in this manner is set large, an output which is approximated to a similar input case can be outputted even for a totally new case. Accordingly, the present embodiment is applicable to various recognitions and diagnosis.

50 A comparison between the calculation by the present invention and the calculation by the conventional back propagation learning will be explained. The comparison was made by measuring corona noises of GIS in a known method, frequency analyzing the data stored in a flexible disk and reading representative data.

In the conventional back propagation learning method, the neural network shown in Table 9 was used. In order to determine a structure of a sample GIS, the number of outputs from the output layer was set to 3. It was intended to determine up to 8 ( $= 2 \times 2 \times 2$ ) types.

EP 0 616 291 A2

Table 9

Number of input data per case	10
Number of outputs of output layer	3
Number of hidden layers	1
Number of neurons of hidden layer	5
Number of cases	14

Three preferable outputs (teacher data) for each case are designated as shown in Table 10.

Table 10

Case Number	Teacher Data		
	Output 1	Output 2	Output 3
1	0	0	0
2	0	0	0
3	1	0	0
4	0	1	0
5	0	0	1
6	0	0	1
7	1	1	0
8	1	1	0
9	1	0	1
10	1	0	0
11	0	1	1
12	0	1	1
13	1	0	1
14	1	0	1

Outputs calculated by the conventional back propagation learning method are shown in Tables 11 and 12.

EP 0 616 291 A2

Table 11

Case Number	After 400 times of learning			After 800 times of learning		
	Output 1	Output 2	Output 3	Output 1	Output 2	Output 3
1	0.000	<u>0.562</u>	0.398	0.000	0.478	<u>0.504</u>
2	0.042	0.431	0.392	0.000	0.133	<u>0.508</u>
3	0.922	0.018	<u>0.575</u>	1.000	0.115	0.160
4	0.000	0.966	<u>0.548</u>	0.000	1.000	0.267
5	0.071	0.163	<u>0.461</u>	0.000	0.000	1.000
6	0.098	0.166	1.000	0.002	0.000	1.000
7	0.957	<u>0.208</u>	0.486	1.000	0.956	0.000
8	0.951	1.000	0.272	1.000	1.000	0.000
9	1.000	0.200	<u>0.394</u>	0.965	0.000	0.980
10	0.975	0.131	0.452	1.000	0.046	0.136
11	0.000	0.981	1.000	0.000	1.000	1.000
12	0.000	0.552	<u>0.398</u>	0.000	0.507	<u>0.465</u>
13	0.996	0.020	0.972	0.973	0.009	1.000
14	0.970	0.031	0.964	0.980	0.019	1.000
Calculation Time	1 hour 47 minutes			3 hours 34 minutes		
Total Error	1.259152			0.456008		

Table 12

Case Number	After 1200 times of learning		
	Output 1	Output 2	Output 3
1	0.000	<u>0.580</u>	<u>0.566</u>
2	0.000	0.248	0.155
3	1.000	0.050	0.027
4	0.000	1.000	0.069
5	0.000	0.000	1.000
6	0.022	0.000	1.000
7	1.000	0.998	0.000
8	1.000	1.000	0.000
9	0.990	0.000	0.994
10	1.000	0.000	0.038
11	0.000	1.000	1.000
12	0.000	0.601	0.537
13	0.984	0.000	1.000
14	0.989	0.005	1.000
Calculation Time	5 hours 21 minutes		
Total Error	0.360491		

In counting the number of times of learning, when the learning is completed one time for each case and all cases have been completed, it is counted as one time of learning. The numbers of times of learning are 400 and 800 in Table 11, and 1200 in Table 12. It is ideal that the calculation output completely matches to Table 10 but it does not completely match in the present example. Where "0" is expected as the output, the number no smaller than 0.5 is not permissible. Similarly, where "1" is expected as the output, the number smaller than 0.5 is not permissible. The impermissible numbers are shown with underscore. As the number of times of learning increases from 400 to 800 and 1200, the number of underscored numbers decreases from 7 to 3 and 2. This shows that the calculation accuracy is enhanced. The times required for the calculation and the sums of the errors of the respective cases at the end of the calculation are also shown. In the present example, it is seen that five hours and twenty minutes are required for the 1200 times of learning and the decrease of the error is slow when the number of time of learning exceeds a certain number. The calculation of the back propagation learning is proceeded while the number of times of learning and the total error are displayed on the CRT. The total error does not decrease monotonously relative to the number of times of learning and it suddenly increases after decrease but gradually decreases as a general trend. In general, Table 12 is still insufficient for the required calculation precision and further learning is required.

Calculation result by the method of Fig. 2 of the present invention by using the same input data is shown in Table 13.

Table 13

Case Number	Calculation Output		
	Output 1	Output 2	Output 3
1	0	0	0
2	0	0	0
3	1	0	0
4	0	1	0
5	0	0	1
6	0	0	1
7	1	1	0
8	1	1	0
9	1	0	1
10	1	0	0
11	0	1	1
12	0	1	1
13	1	0	1
14	1	0	1
Calculation Time	0.84 second		
Total Error	0.000		

As seen from Table 13, the calculation output rounded at fourth digit below a decimal point completely matches to the teacher data shown in Table 10. The time required for the calculation of the coupling weights is 0.84 second.

In the method of the present invention, since the analytical calculation is made and the iterative calculation to converge the number is not included, the calculation time is very short.

The present embodiment offers the following advantages:

(1) The number of hidden layers of the neural network and the number of neurons thereof can be determined base on the input data and the teacher data, and the coupling weights can be analytically calculated. Accordingly, the prior art reverse propagation iterative learning which requires a long time is not necessary and the time from the data input to the recognition and diagnosis can be reduced.

(2) Some time of trial and error required in the prior art method to select the number of hidden layers and the number of neurons thereof is not necessary. Accordingly, the calculation can be proceeded efficiently.

(3) It is applicable to a large scale neural network having a large number of input data and teacher data.

(4) Where the input data is added together with the teacher data, whether it is worthwhile for adding can be determined, and if it is determined worthless, the addition can be cancelled to save the memory space of the computer.

(5) Since the iterative learning is not necessary, the calculation of the coupling weights between neuron to the diagnosis can be done by more economic computer such as a notebook personal computer.

(6) Where the teaching of the present invention is applied to the neural network which requires the conventional back propagation method, the time required for the learning is significantly reduced because the number of hidden layers and the number of neurons are known, the relation between the coupling weights are known and the number of coupling weights which actually require the learning is very small, and the functions of the conventional program can be utilized as they are.

(7) Where the number of case of input data is large, they are classified into a plurality of classes based on some features of the cases and the learning and diagnosis are made for each class. Accordingly, the precision of diagnosis is enhanced even for similar or confusing cases.

(8) A correct solution is attained for the input data having a scatter if the scatter is within the predetermined range of scatter.

(9) A good method for selecting the coupling weights of the initial stage to classify the input data is provided and a selection efficiency is enhanced.

5 (10) In classifying the input data, a portion of data is not used directly but reciprocals thereof are used so that the input data can be efficiently classified.

(11) For the unlearned data, it may be inferred from the relation with adjacent data as it is done in the prior art, or the fact of not learnt may be indicated by an appropriate expression such as "?" or a deviation from the learnt data may be indicated so that more correct diagnosis is attained.

10 (12) Where there is teacher data which is a correct solution to the diagnosed input data, the diagnosed input data and the correct solution teacher data are efficiently added to the learning data. Accordingly, the precision of diagnosis is enhanced.

(13) The diagnosis data is inputted, the output is calculated and the most similar learning case is precisely inferred based on the output in a logically correct manner.

15 While the above description is primarily based on a method, the present invention may be implemented by hardware having a similar function. The fields of application of the present invention are not only engineering fields but also biotechnology and medical fields as well as the recognition and diagnosis of physical, chemical and biological phenomena and the recognition and diagnosis in agricultural, educational, economical and other various fields.

20 In accordance with the present invention, the coupling weights can be precisely and readily determined in a system such as the neural network, and the number of hidden layers and the number of neurons can be readily determined.

#### Claims

25

1. A recognition or diagnosis method using a neural network having an input layer, at least a hidden layer and an output layer comprising the steps of:  
 classifying learning data inputted to said input layer based on features thereof;  
 inputting learning data for each class;  
 30 calculating a total sum for each case;  
 correlating corresponding teacher data to the total sum in the order of magnitude for each output of said output layer; and  
 determining coupling weights based on the teacher data corresponding to a case having a maximum total sum, total sums before and after the change of the teacher data and the teacher data.

35

2. A recognition or diagnosis method using a neural network having an input layer, at least a hidden layer and an output layer comprising the steps of:  
 inputting input data for each value for each case based on an upper limit and a lower limit of the input data of each case inputted to said input layer;  
 40 calculating a total sum for each case;  
 correlating corresponding teacher data to the total sum in the order of magnitude for each output of said output layer; and  
 determining coupling weights based on the teacher data corresponding to a case having a maximum total sum, total sums before and after the change of the teacher data and the teacher data.

45

3. A recognition or diagnosis method according to Claim 2 wherein the input data inputted to said input layer are products of the upper limit and the lower limit multiplied by the same coupling weight.

50

4. A recognition or diagnosis method using a neural network having an input layer, at least a hidden layer and an output layer comprising the steps of:  
 selecting coupling weights such that an input sum determined for a case other than a particular case inputted to said input layer does not fall between two input sums determined based on an upper limit and a lower limit of input data of said particular case and determining the input sum;  
 inputting said input sum as input data;  
 55 calculating an input sum for each case;  
 calculating a total sum for each case;  
 correlating corresponding teacher data to the input sum in the order of magnitude for each output of said output layer; and

determining coupling weights based on the teacher data corresponding to a case having a maximum total sum, total sums before and after the change of the teacher data and the teacher data.

- 5 5. A recognition or diagnosis method according to Claim 4 wherein the teacher data of said particular case and the teacher data of a case other than said particular case are different from each other.
- 10 6. A recognition or diagnosis method according to Claim 1 wherein two hidden layers are provided and the coupling weights are selected such that when the teacher data corresponding to the case having the maximum total sum is 0, the second hidden layer functions as an AND circuit and said output layer functions as an OR circuit.
- 15 7. A recognition or diagnosis method according to Claim 1 wherein two hidden layers are provided and the coupling weights are selected such that when the teacher data corresponding to the case having the maximum total sum is 1, the second hidden layer functions as an OR circuit and said output layer functions as an AND circuit.
- 20 8. A recognition or diagnosis method according to Claim 1 wherein said data are used by reciprocals thereof for each case and a reversal operation in the magnitude of the input data is executed.
- 25 9. A recognition or diagnosis method according to Claim 1 wherein a back propagation iterative learning method is used to calculate the coupling weights which are principal parts of the neural network.
- 30 10. A recognition or diagnosis method using a neural network having an input layer, at least a hidden layer and an output layer comprising the steps of:  
inputting learning data to said input layer for each case;  
calculating a total sum for each case;  
correlating corresponding teacher data to the total sum in the order of magnitude for each output of said output layer;  
determining coupling weights based on the teacher data corresponding to a case having a maximum total sum, total sums before and after the change of the teacher data and the teacher data;  
and  
indicating the deviation from a specified range of scatter of the learning data when input diagnosis data is out of the range.
- 35 11. A recognition or diagnosis method using a neural network having an input layer, at least a hidden layer and an output layer comprising the steps of:  
inputting learning data to said input layer for each case;  
calculating a total sum for each case;  
correlating corresponding teacher data to the total sum in the order of magnitude for each output of said output layer;  
40 determining coupling weights based on the teacher data corresponding to a case having a maximum total sum, total sums before and after the change of the teacher data and the teacher data;  
and  
transferring diagnosis data and the corresponding correct solution as a portion of the learning data after the diagnosis when the correct solution to the diagnosis data is known in inputting the diagnosis data.  
45
- 50 12. A recognition or diagnosis method using a neural network having an input layer, at least a hidden layer and an output layer comprising the steps of:  
inputting learning data to said input layer for each case;  
calculating a total sum for each case;  
correlating corresponding teacher data to the total sum in the order of magnitude for each output of said output layer;  
55 determining coupling weights based on the teacher data corresponding to a case having a maximum total sum, total sums before and after the change of the teacher data and the teacher data;  
and  
calculating an output for diagnosis data when the diagnosis data is inputted;  
inferring at least one probable learning data from the output;



comparing the diagnosis data with the learning data; and  
indicating the result thereof.

- 5 13. A recognition or diagnosis method according to Claim 12 wherein said inferred learning data includes a scatter and the diagnosis data is compared with the learning data having the scatter and the result thereof is indicated.
- 10 14. A recognition or diagnosis method according to Claim 1 wherein a first coupling weight to be multiplied to the input data of said input layer is selected to be different from other coupling weights.
- 15 15. A recognition or diagnosis method comprising the steps of:  
comparing diagnosis data directly with learning data;  
determining whether the diagnosis data and the learning data of a particular case totally match,  
partially match or totally mismatch; and  
indicating the result thereof.
- 20 16. A recognition or diagnosis method comprising the steps of:  
comparing diagnosis data directly with learning data having a range of scatter;  
determining whether the diagnosis data is totally within said range, partially within said range or  
totally out of said range; and  
indicating the result thereof.
- 25 17. A recognition or diagnosis method comprising the steps of:  
preparing simultaneous equations by using calculation formulas of inputs to a neuron based on an  
upper limit and a lower limit of the input data having a range of scatter;  
solving the simultaneous equations to determine coupling weights; and  
utilizing a neural network based on the coupling weights.
- 30 18. A recognition or diagnosis method by using a neural network having an input layer, at least a hidden  
layer and an output layer weighted by coupling weights comprising the step of:  
self-evaluating teacher data for determining the coupling weights.
- 35 19. An apparatus for diagnosing an electrical circuit to be used under a plurality of conditions by using a  
neural network comprising means for taking the plurality of conditions as learning data to determine  
coupling weights which form said neural network.

40

45

50

55

FIG. 1

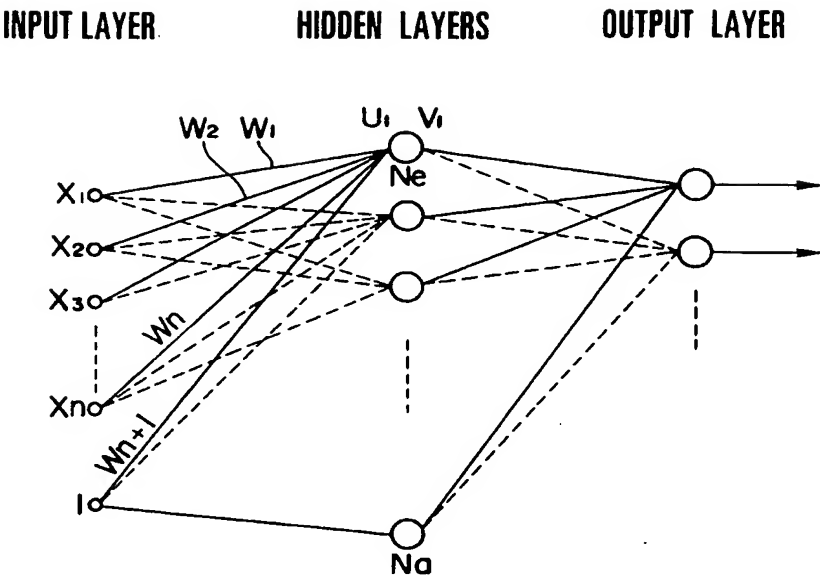


FIG.2

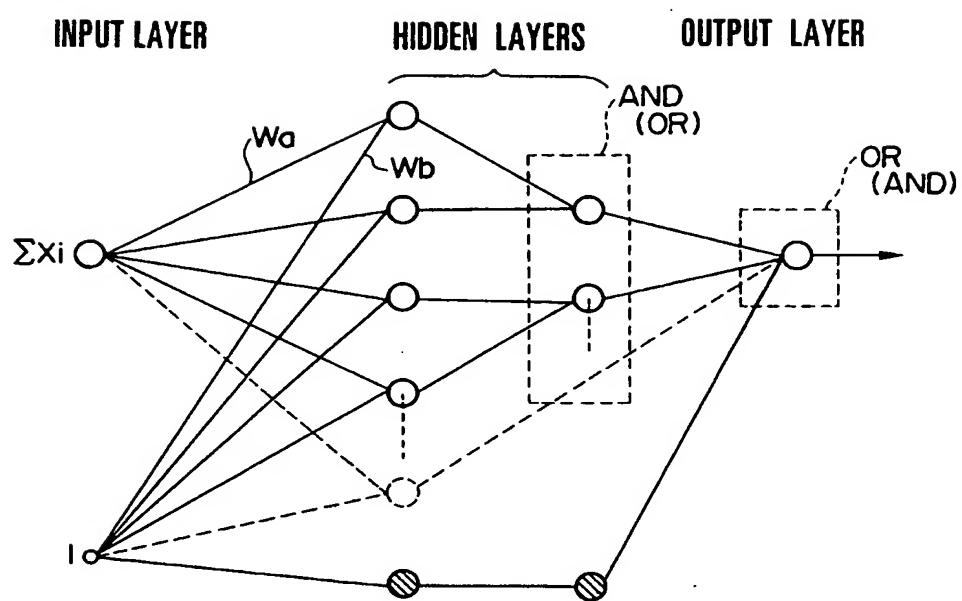


FIG.3

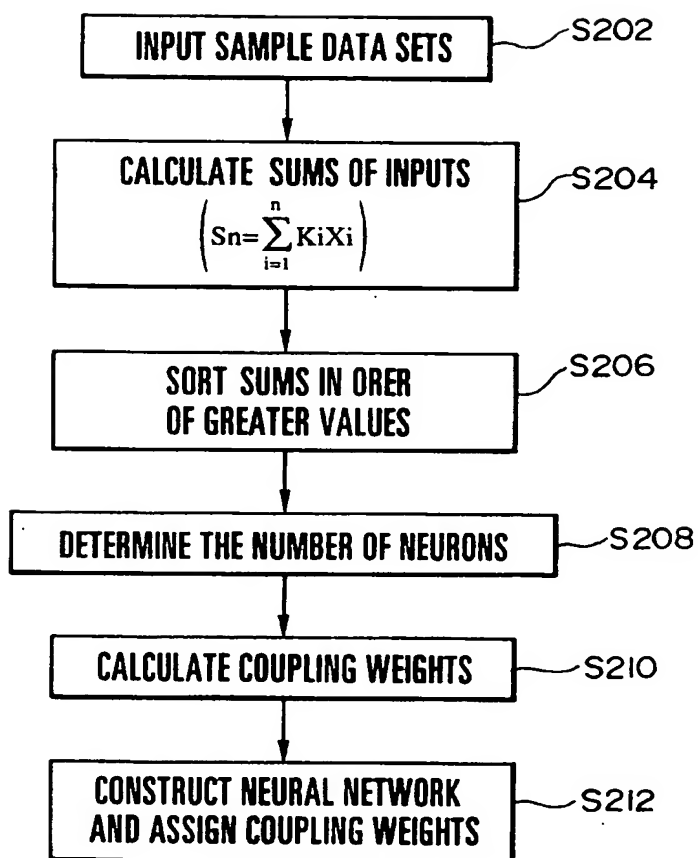


FIG.4

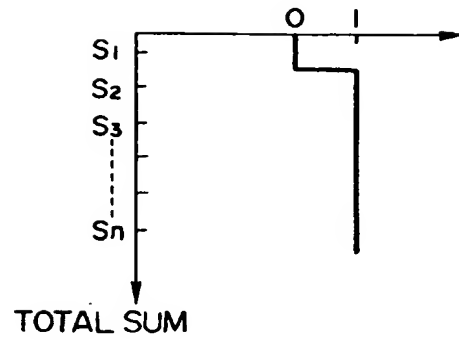


FIG.5

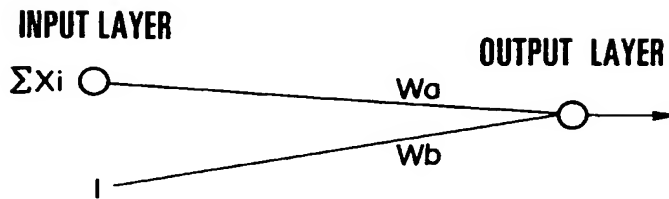


FIG.6

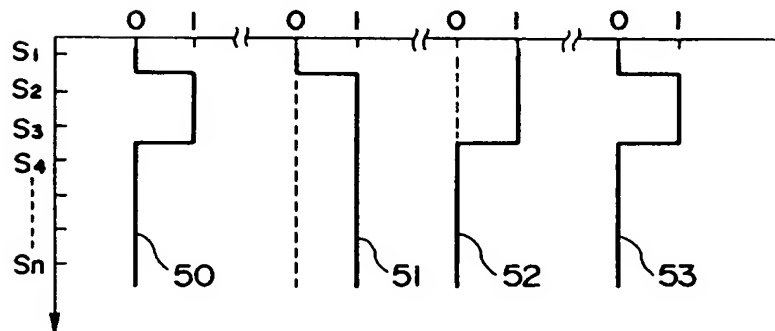


FIG.7

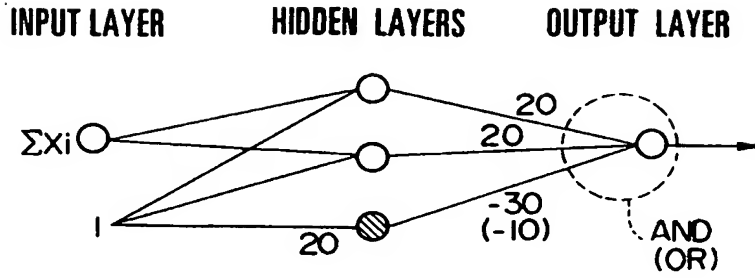


FIG.8

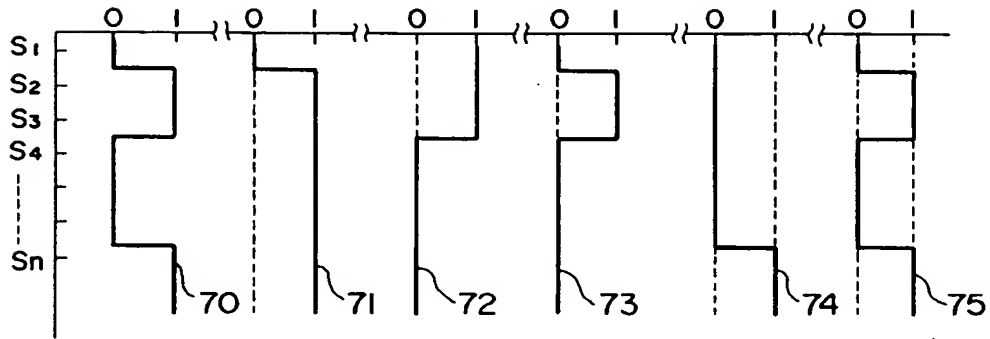


FIG.9

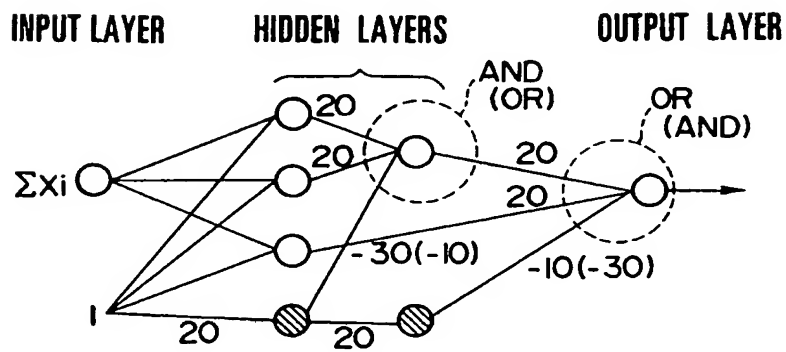


FIG.10

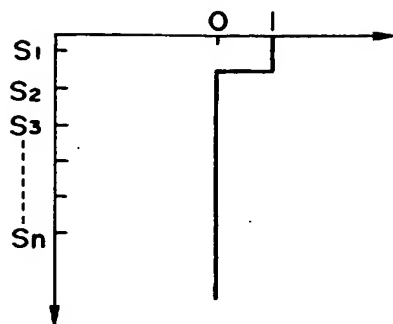


FIG.11

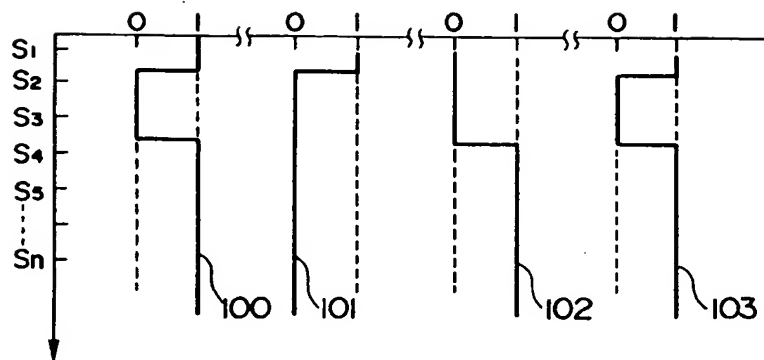


FIG.12

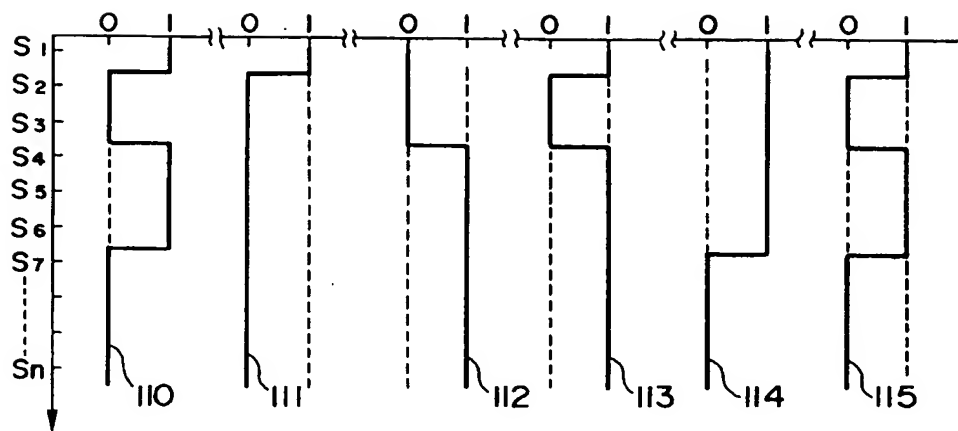
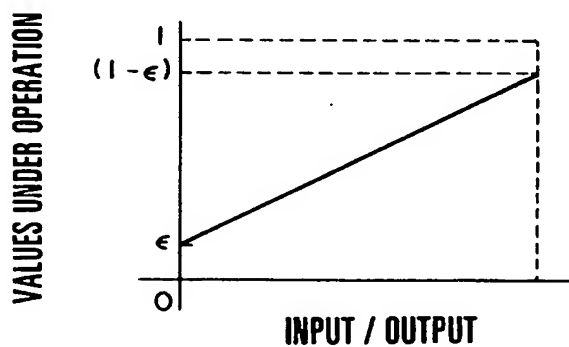


FIG.13





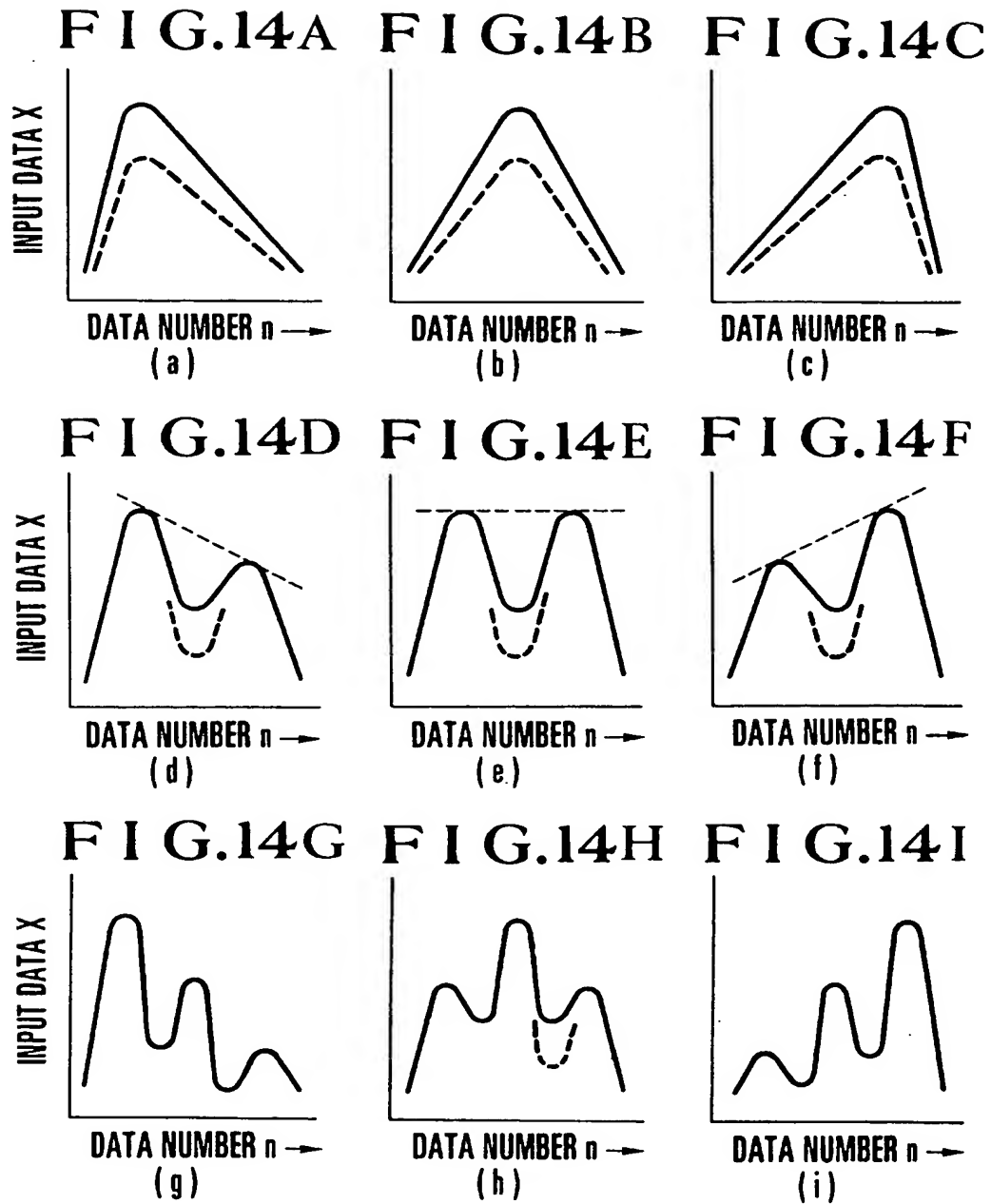


FIG.15

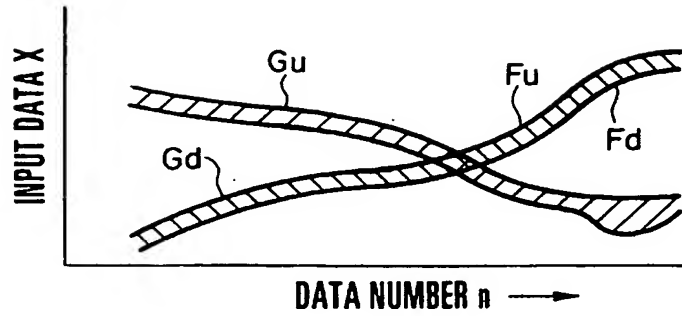


FIG.16

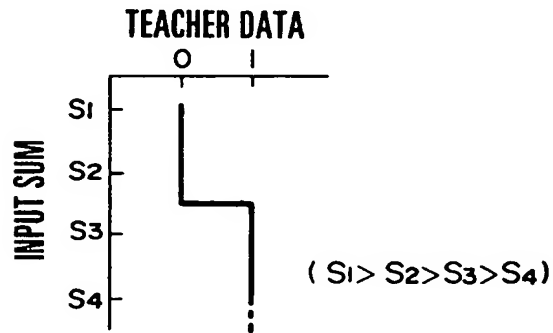


FIG.17

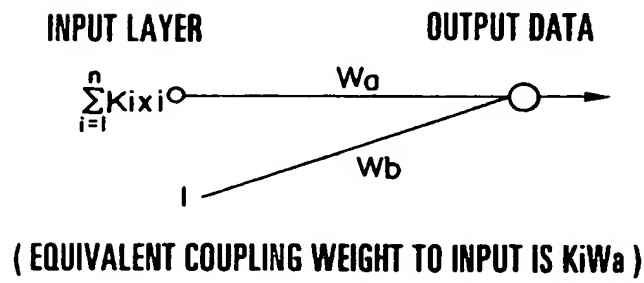


FIG.18

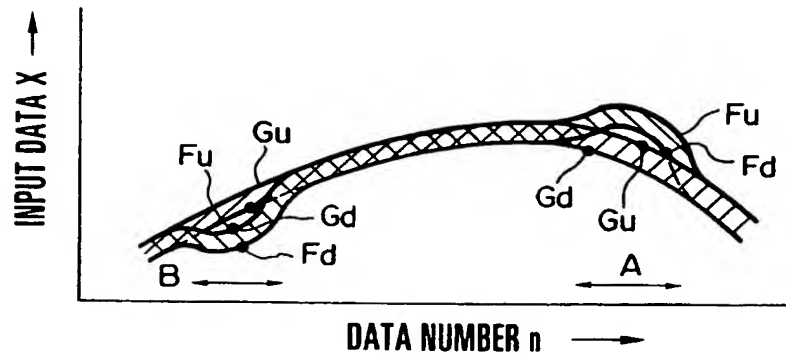


FIG.19

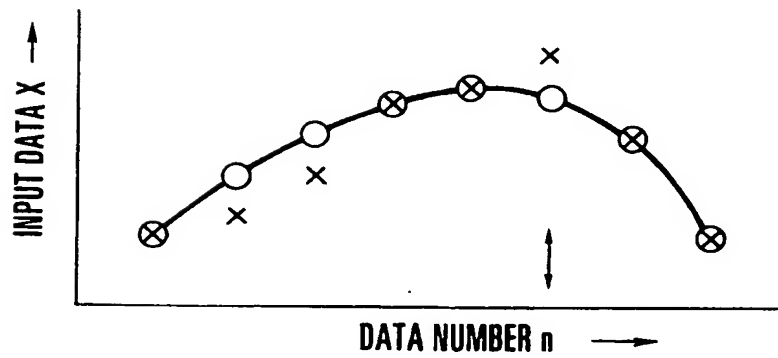


FIG.20

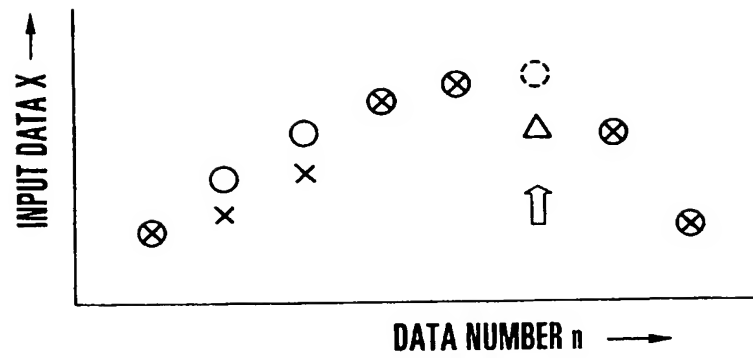


FIG.21

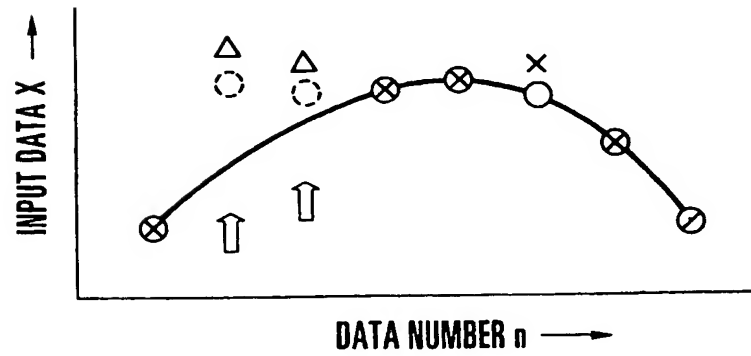


FIG.22

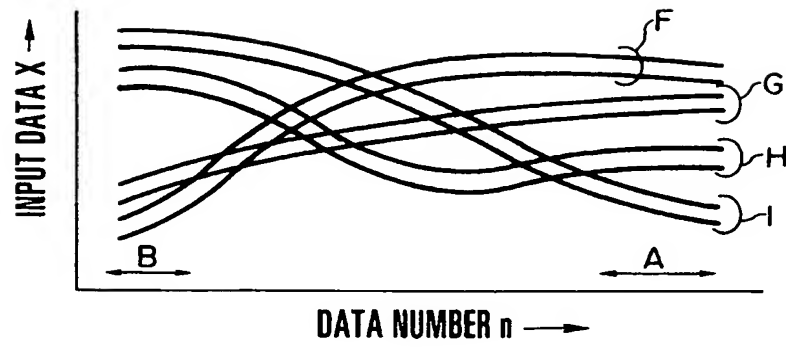
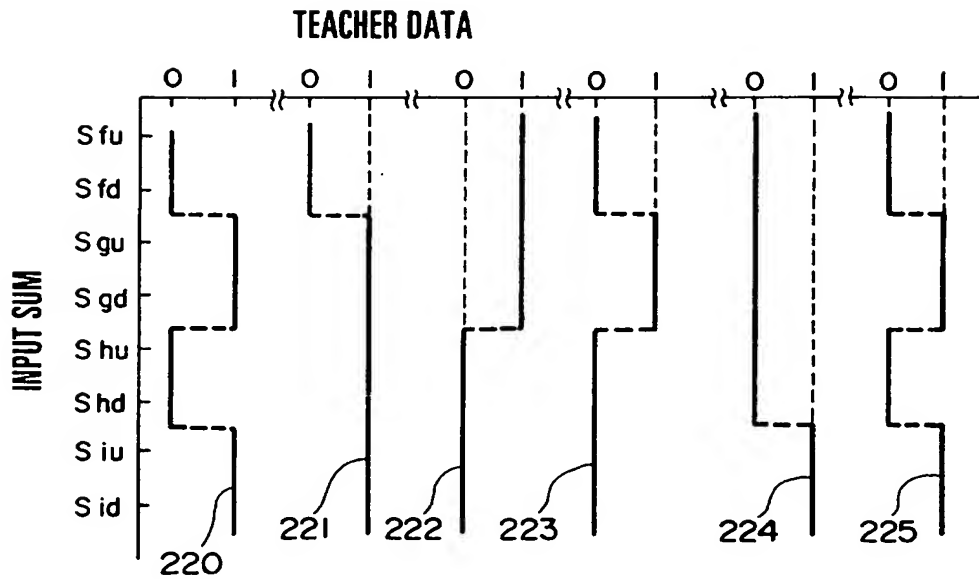
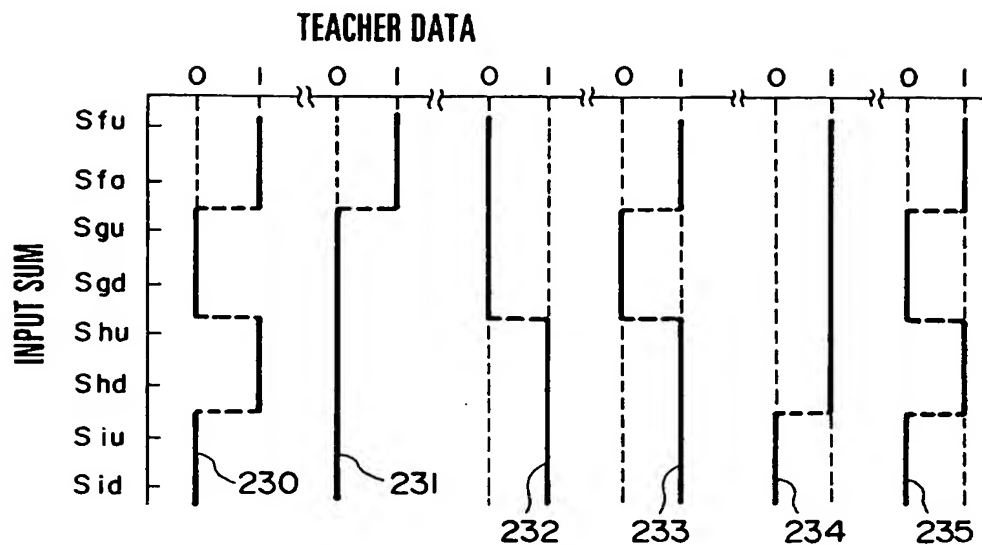


FIG.23



**FIG.24**



**FIG. 25**

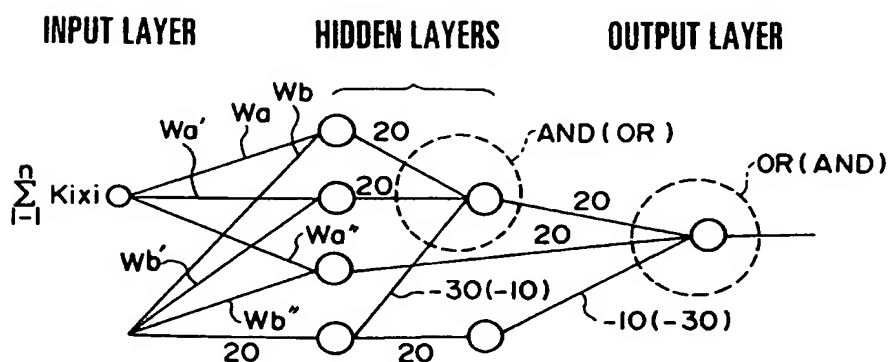


FIG. 26

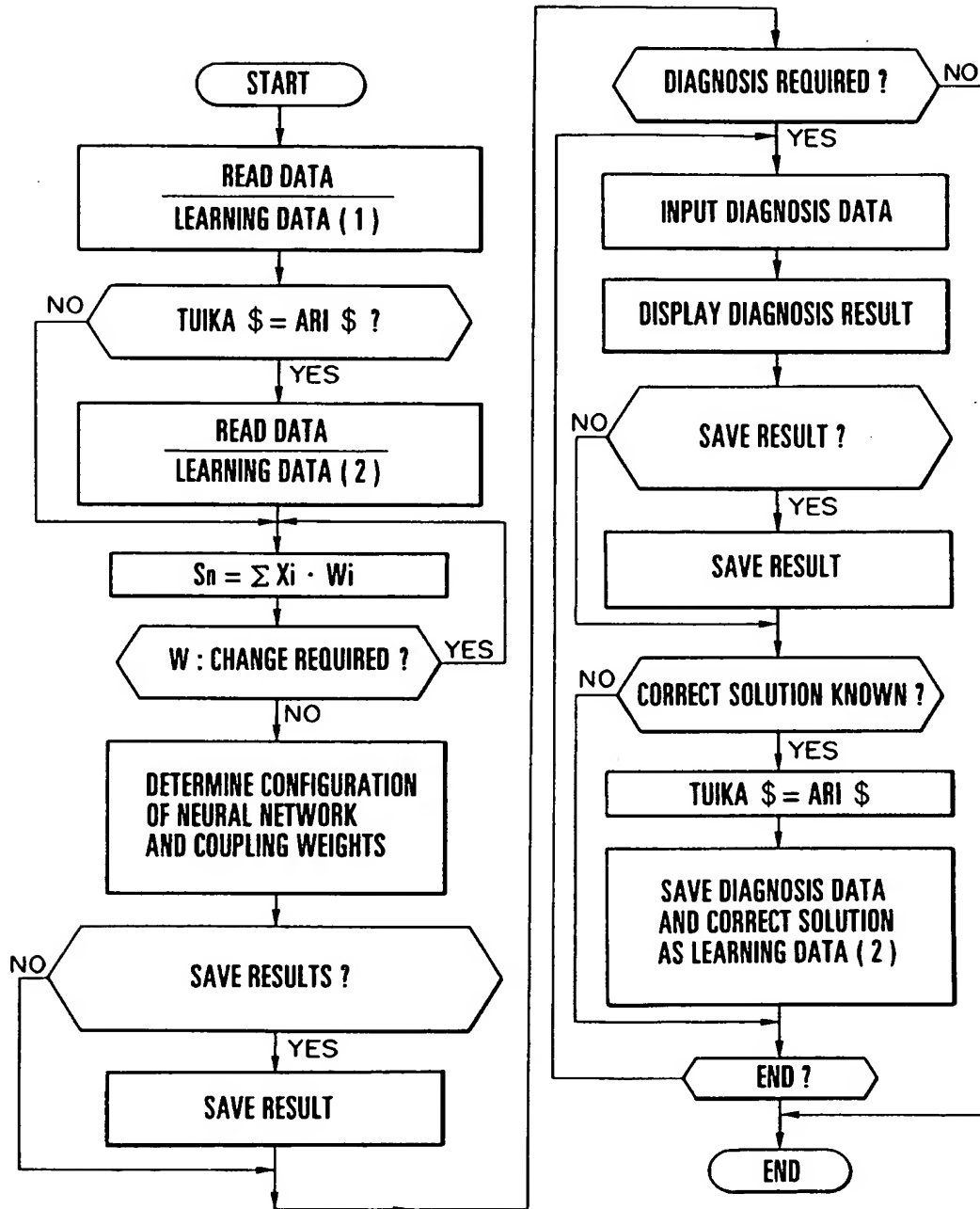


FIG. 27

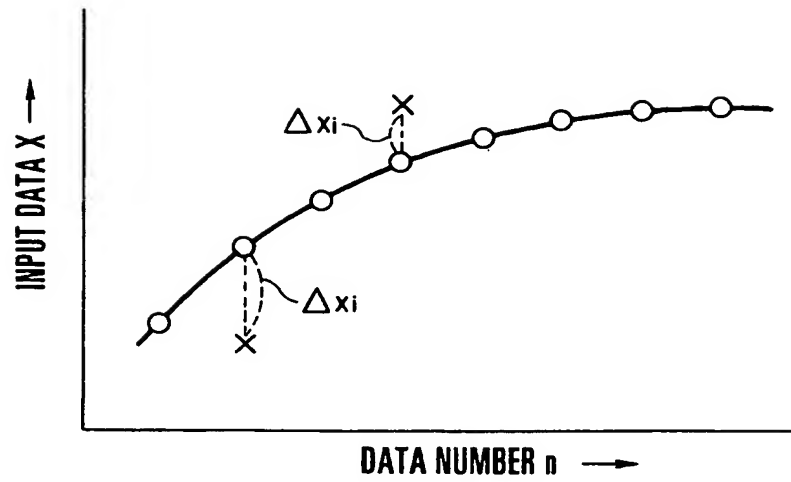


FIG. 28

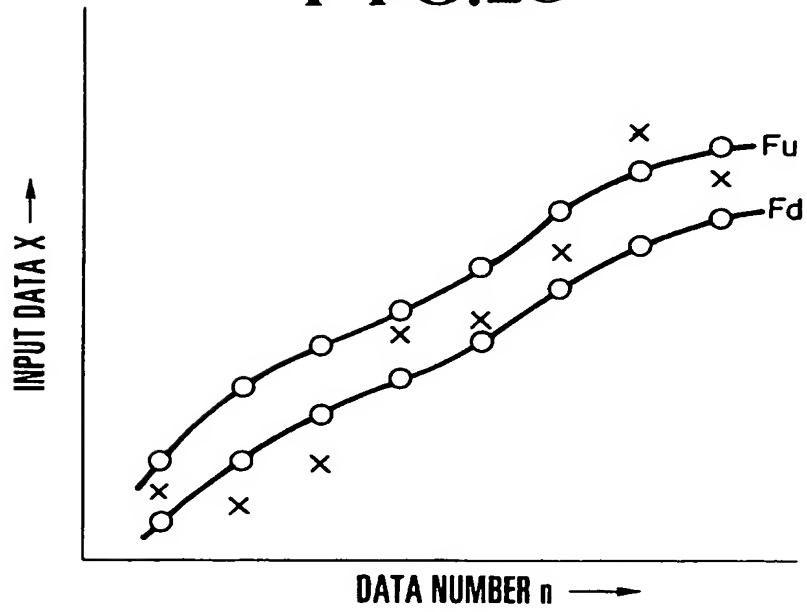
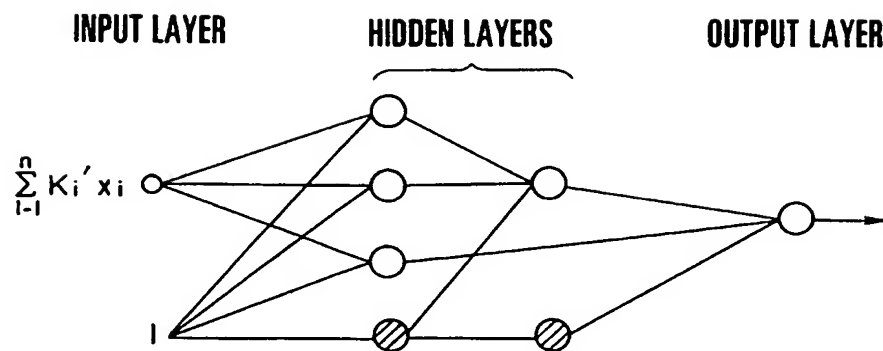
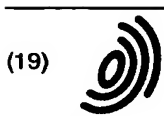




FIG.29





(19)

Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 0 616 291 A3

(12)

## EUROPEAN PATENT APPLICATION

(88) Date of publication A3:  
10.07.1996 Bulletin 1996/28

(51) Int. Cl.<sup>6</sup>: G06F 15/80

(43) Date of publication A2:  
21.09.1994 Bulletin 1994/38

(21) Application number: 94101489.6

(22) Date of filing: 01.02.1994

(84) Designated Contracting States:  
CH DE FR LI SE

(30) Priority: 18.03.1993 JP 58328/93

(71) Applicant: HITACHI, LTD.  
Chiyoda-ku, Tokyo 100 (JP)

(72) Inventors:  
• Takahashi, Isao  
Maebashi-shi (JP)

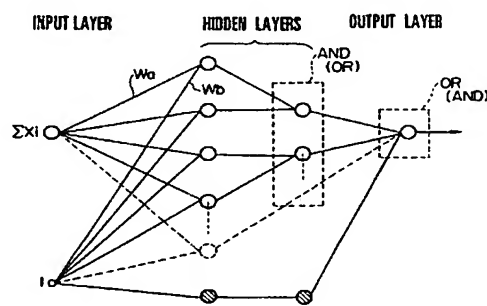
• Endo, Fumihiro  
Hitachi-shi (JP)  
• Yamagiwa, Tokio  
Hitachi-shi (JP)

(74) Representative: Strehl Schübel-Hopf Groening &  
Partner  
Maximilianstrasse 54  
80538 München (DE)

(54) A method of configuring a neural network and a diagnosis/recognition system using the same

(57) In a recognition/diagnosis method, a relation between total sums of inputs of respective cases and teacher data is listed in the order of magnitude of the total sums of inputs. Based on the value of the teacher data for the input having the maximum total sum and the number of times of change of teacher data in the table are considered, a configuration of a neural network (the number of hidden layers and the number of neurons thereof) is determined. Coupling weights are analytically calculated based on the table.

FIG.2



EP 0 616 291 A3



European Patent  
Office

## EUROPEAN SEARCH REPORT

Application Number  
EP 94 10 1489

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.5)
X	IEEE TRANSACTIONS ON NEURAL NETWORKS, vol.2, no.4, 1 July 1991 pages 467 - 471, XP000228662 SARTORI M A ET AL 'A SIMPLE METHOD TO DERIVE BOUNDS ON THE SIZE AND TO TRAIN MULTILAYER NEURAL NETWORKS'	17,18	G06F15/80
A	* the whole document *	1,2,4, 10-12	
A	--- WO-A-90 07158 (DU PONT) 28 June 1990  * page 5, line 4 - page 6, line 9 *	1,2,4, 10-12, 17,18	
A	--- PROCEEDINGS OF THE INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORK (IJCNN), BALTIMORE, JUNE 7 - 11, 1992, vol.1, 7 June 1992, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS pages 676 - 681, XP000340280 DAVIS D T ET AL 'ATTENTIONAL FOCUS TRAINING BY BOUNDARY REGION DATA SELECTION'	1,2,4, 10-12	
	* page 676, line 1 - page 677, line 22 *		TECHNICAL FIELDS SEARCHED (Int.Cl.5) G06F
A	--- SPEECH PROCESSING 2, AUDIO, NEURAL NETWORKS, UNDERWATER ACOUSTICS, SAN FRANCISCO, MAR. 23 - 26, 1992, vol.2, 23 March 1992, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS pages 401 - 404, XP000357022 RE R 'A TECHNIQUE FOR DEFINING THE ARCHITECTURE AND WEIGHTS OF A NEURAL IMAGE CLASSIFIER'	1	
	* page 401, left column, line 1 - page 402, right column, line 54 *		
	--- -/--		
The present search report has been drawn up for all claims			
Place of search THE HAGUE		Date of completion of the search 22 February 1996	Examiner SCHENKELS P.F.
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document		T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons ----- & : member of the same patent family, corresponding document	

EPO FORM 1501 (12/94) (P01C01)



European Patent  
Office

# EUROPEAN SEARCH REPORT

Application Number  
EP 94 10 1489

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.5)
A	<p>PROCEEDINGS OF THE INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORK, SINGAPORE, NOV. 18 - 21, 1991, vol.1 OF 3, 18 November 1991, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS pages 813 - 820, XP000325724</p> <p>CHU Y C ET AL 'A NEURAL NETWORK WHICH LEARNS DECISION BOUNDARIES WITH NONLINEAR CLUSTERING'</p> <p>* page 813, line 1 - line 35 *</p> <p>-----</p>	1	
			TECHNICAL FIELDS SEARCHED (Int.Cl.5)
The present search report has been drawn up for all claims			
Place of search <b>THE HAGUE</b>		Date of completion of the search <b>22 February 1996</b>	Examiner <b>SCHENKELS P.F.</b>
<p><b>CATEGORY OF CITED DOCUMENTS</b></p> <p>X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document</p> <p>T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons</p> <p>----- &amp; : member of the same patent family, corresponding document</p>			

EPO FORM 1503 (12.12.1994) (PM/CI)



European Patent  
Office

### CLAIMS INCURRING FEES

The present European patent application comprised at the time of filing more than ten claims.

- ☐ All claims fees have been paid within the prescribed time limit. The present European search report has been drawn up for all claims.
- ☐ Only part of the claims fees have been paid within the prescribed time limit. The present European search report has been drawn up for the first ten claims and for those claims for which claims fees have been paid, namely claims:
- ☐ No claims fees have been paid within the prescribed time limit. The present European search report has been drawn up for the first ten claims.

### LACK OF UNITY OF INVENTION

The Search Division considers that the present European patent application does not comply with the requirement of unity of invention and relates to several inventions or groups of inventions, namely:

See Sheet B.

- ☐ All further search fees have been paid within the fixed time limit. The present European search report has been drawn up for all claims.
- ☐ Only part of the further search fees have been paid within the fixed time limit. The present European search report has been drawn up for those parts of the European patent application which relate to the inventions in respect of which search fees have been paid, namely claims:
- ☒ None of the further search fees has been paid within the fixed time limit. The present European search report has been drawn up for those parts of the European patent application which relate to the invention first mentioned in the claims, namely claims: 1-14, 17, 18



European Patent  
Office

EP 94101489.6 - B -

**LACK OF UNITY OF INVENTION**

The Search Division considers that the present European patent application does not comply with the requirement of unity of invention and relates to several inventions or groups of inventions, namely:

- 1: Claims 1-14, 17, 18:  
Neural network with weight calculation
- 2: Claims 15, 16:  
A recognition or diagnosis method by data comparison
- 3: Claim 19  
Diagnosis apparatus that uses conditions as learning data